

**PCT**

WORLD INTELLECTUAL PROPERTY ORGANIZATION  
International Bureau



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

<b>(51) International Patent Classification <sup>6</sup> :</b> <b>C07H 21/02, C07K 1/00, C12N 15/00, 1/20, C12P 21/06</b>	<b>A1</b>	<b>(11) International Publication Number: WO 97/02280</b> <b>(43) International Publication Date: 23 January 1997 (23.01.97)</b>
<b>(21) International Application Number:</b> PCT/US95/08295 <b>(22) International Filing Date:</b> 30 June 1995 (30.06.95) <b>(71) Applicant:</b> HUMAN GENOME SCIENCES, INC. [US/US]; 9410 Key West Avenue, Rockville, MD 20850-3338 (US). <b>(72) Inventors:</b> JI, Hongjun; 13144 Country Ridge Drive, German- town, MD 20874 (US). ROSEN, Craig, A.; 22400 Rolling Hill Road, Laytonsville, MD 20850-3338 (US). <b>(74) Agents:</b> OLSTEIN, Elliot, M. et al.; Carella, Byrne, Bain, Gilfillan, Cecchi, Stewart & Olstein, 6 Becker Farm Road, Roseland, NJ 07068 (US).		<b>(81) Designated States:</b> AU, CA, CN, JP, KR, MX, SI, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).  <b>Published</b> <i>With international search report.</i>
<b>(54) Title:</b> BREAST SPECIFIC GENES AND PROTEINS  <b>(57) Abstract</b>  Human breast specific gene polypeptides and DNA (RNA) encoding such polypeptides and a procedure for producing such polypeptides by recombinant techniques is disclosed. Also disclosed are methods for utilizing such polynucleotides or polypeptides as a diagnostic marker for breast cancer and as an agent to determine if breast cancer has metastasized. Also disclosed are antibodies specific to the breast specific gene polypeptides which may be used to target cancer cells and be used as part of a breast cancer vaccine. Methods of screening for antagonists for the polypeptide and therapeutic uses thereof are also disclosed.		

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AM	Armenia	GB	United Kingdom	MW	Malawi
AT	Austria	GE	Georgia	MX	Mexico
AU	Australia	GN	Guinea	NE	Niger
BB	Barbados	GR	Greece	NL	Netherlands
BE	Belgium	HU	Hungary	NO	Norway
BF	Burkina Faso	IE	Ireland	NZ	New Zealand
BG	Bulgaria	IT	Italy	PL	Poland
BJ	Benin	JP	Japan	PT	Portugal
BR	Brazil	KE	Kenya	RO	Romania
BY	Belarus	KG	Kyrgyzstan	RU	Russian Federation
CA	Canada	KP	Democratic People's Republic of Korea	SD	Sudan
CF	Central African Republic	KR	Republic of Korea	SE	Sweden
CG	Congo	KZ	Kazakhstan	SG	Singapore
CH	Switzerland	LI	Liechtenstein	SI	Slovenia
CI	Côte d'Ivoire	LK	Sri Lanka	SK	Slovakia
CM	Cameroon	LR	Liberia	SN	Senegal
CN	China	LT	Lithuania	SZ	Swaziland
CS	Czechoslovakia	LU	Luxembourg	TD	Chad
CZ	Czech Republic	LV	Latvia	TG	Togo
DE	Germany	MC	Monaco	TJ	Tajikistan
DK	Denmark	MD	Republic of Moldova	TT	Trinidad and Tobago
EE	Estonia	MG	Madagascar	UA	Ukraine
ES	Spain	ML	Mali	UG	Uganda
FI	Finland	MN	Mongolia	US	United States of America
FR	France	MR	Mauritania	UZ	Uzbekistan
GA	Gabon			VN	Viet Nam

### BREAST SPECIFIC GENES AND PROTEINS

This invention relates to newly identified polynucleotides, polypeptides encoded by such polynucleotides, and the use of such polynucleotides and polypeptides for detecting disorders of the breast, particularly the presence of breast cancer and breast cancer metastases. The present invention further relates to inhibiting the production and function of the polypeptides of the present invention. The twenty breast specific genes of the present invention are sometimes hereinafter referred to as "BSG1", "BSG2" etc.

The mammary gland is subject to a variety of disorders that should be readily detectable. Detection may be accomplished by inspection which usually consists of palpation. Unfortunately, so few periodic self-examinations are made that many breast masses are discovered only by accidental palpation. Aspiration of suspected cysts with a fine-gauge needle is another fairly common diagnostic practice. Mammography or xeroradiography (soft-tissue x-ray) of the breast of yet another. A biopsy of a lesion or suspected area is an extreme method of diagnostic test.

There are many types of tumors and cysts which affect the mammary gland. Fibroadenomas is the most common benign breast tumor. As a pathological entity, it ranks third

behind cystic disease and carcinoma, respectively. These tumors are seen most frequently in young people and are usually readily recognized because they feel encapsulated. Fibrocystic disease, a benign condition, is the most common disease of the female breast, occurring in about 20% of pre-menopausal women. Lipomas of the breast are also common and they are benign in nature. Carcinoma of the breast is the most common malignant condition among women and carries with it the highest fatality rate of all cancers affecting this sex. At some during her life, one of every 15 women in the USA will develop cancer of the breast. Its reported annual incidence is 70 per 100,000 females in the population in 1947, rising to 72.5 in 1969 for whites, and rising from 47.8 to 60.1 for blacks. The annual mortality rate from 1930 to the present has remained fairly constant, at approximately 23 per 100,000 female population. Breast cancer is rare in men, but when it does occur, it usually not recognized until late, and thus the results of treatment are poor. In women, carcinoma of the breast is rarely seen before age 30 and the incidence rises rapidly after menopause. For this reason, post-menopausal breast masses should be considered cancer until proved otherwise.

In accordance with an aspect of the present invention, there are provided nucleic acid probes comprising nucleic acid molecules of sufficient length to specifically hybridize to the RNA transcribed from the human breast specific genes of the present invention or to DNA corresponding to such RNA.

In accordance with another aspect of the present invention there is provided a method of and products for diagnosing breast cancer formation and breast cancer metastases by detecting the presence of RNA transcribed from the human breast specific genes of the present invention or DNA corresponding to such RNA in a sample derived from a host.

In accordance with yet another aspect of the present invention, there is provided a method of and products for diagnosing breast cancer formation and breast cancer metastases by detecting an altered level of a polypeptide corresponding to the breast specific genes of the present invention in a sample derived from a host, whereby an elevated level of the polypeptide indicates a breast cancer diagnosis.

In accordance with another aspect of the present invention, there are provided isolated polynucleotides encoding human breast specific polypeptides, including mRNAs, DNAs, cDNAs, genomic DNAs, as well as antisense analogs and biologically active and diagnostically or therapeutically useful fragments thereof.

In accordance with still another aspect of the present invention there are provided human breast specific genes which include polynucleotides as set forth in the sequence listing.

In accordance with a further aspect of the present invention, there are provided novel polypeptides encoded by the polynucleotides, as well as biologically active and diagnostically or therapeutically useful fragments, analogs and derivatives thereof.

In accordance with yet a further aspect of the present invention, there is provided a process for producing such polypeptides by recombinant techniques comprising culturing recombinant prokaryotic and/or eukaryotic host cells, containing a polynucleotide of the present invention, under conditions promoting expression of said proteins and subsequent recovery of said proteins.

In accordance with yet a further aspect of the present invention, there are provided antibodies specific to such polypeptides, which may be employed to detect breast cancer cells or breast cancer metastasis.

In accordance with another aspect of the present invention, there are provided processes for using one or more of the polypeptides of the present invention to treat breast cancer and for using the polypeptides to screen for compounds which interact with the polypeptides, for example, compounds which inhibit or activate the polypeptides of the present invention.

In accordance with yet another aspect of the present invention, there is provided a screen for detecting compounds which inhibit activation of one or more of the polynucleotides and/or polypeptides of the present invention which may be used to therapeutically, for example, in the treatment of breast cancer.

In accordance with yet a further aspect of the present invention, there are provided processes for utilizing such polypeptides, or polynucleotides encoding such polypeptides, for in vitro purposes related to scientific research, synthesis of DNA and manufacture of DNA vectors.

These and other aspects of the present invention should be apparent to those skilled in the art from the teachings herein.

The following drawings are illustrative of embodiments of the invention and are not meant to limit the scope of the invention as encompassed by the claims.

Figure 1 is a full length cDNA sequence of breast specific gene 1 of the present invention.

Figure 2 is a partial cDNA sequence and the corresponding deduced amino acid sequence of breast specific gene 2 of the present invention.

Figure 3 is a partial cDNA sequence and deduced amino acid sequence of breast specific gene 3 of the invention.

Figure 4 is a partial cDNA sequence and the corresponding deduced amino acid sequence of breast specific gene 4 of the present invention.

Figure 5 is a partial cDNA sequence of breast specific gene 5 of the present invention.

Figure 6 is a partial cDNA and deduced amino acid sequence of breast specific gene 6 of the present invention.

Figure 7 is a partial cDNA sequence of breast specific gene 7 of the present invention.

Figure 8 is a partial cDNA sequence of breast specific gene 8 of the present invention.

Figure 9 is a partial cDNA sequence of breast specific gene 9 of the present invention.

Figure 10 is a partial cDNA sequence of breast specific gene 10 of the present invention.

Figure 11 is a partial cDNA sequence of breast specific gene 11 of the present invention.

Figure 12 is a partial cDNA sequence of breast specific gene 12 of the present invention.

Figure 13 is a partial cDNA sequence of breast specific gene 13 of the present invention.

Figure 14 is a partial cDNA sequence of breast specific gene 14 of the present invention.

Figure 15 is a partial cDNA sequence of breast specific gene 15 of the present invention.

Figure 16 is a partial cDNA sequence of breast specific gene 16 of the present invention.

Figure 17 is a partial cDNA sequence of breast specific gene 17 of the present invention.

Figure 18 is a partial cDNA sequence of breast specific gene 18 of the present invention.

Figure 19 is a partial cDNA sequence of breast specific gene 19 of the present invention.

Figure 20 is a partial cDNA sequence of breast specific gene 20 of the present invention.

The term "breast specific gene" means that such gene is primarily expressed in tissues derived from the breast, and such genes may be expressed in cells derived from tissues

other than from the breast. However, the expression of such genes is significantly higher in tissues derived from the breast than from non-breast tissues.

In accordance with one aspect of the present invention there is provided a polynucleotide which encodes the mature polypeptides having the deduced amino acid sequence of Figure 1 (SEQ ID NO:1) and fragments, analogues and derivatives thereof.

In accordance with a further aspect of the present invention there is provided a polynucleotide which encodes the same mature polypeptide as a human gene having a coding portion which contains a polynucleotide which is at least 90% identical (preferably at least 95% identical and most preferably at least 97% or 100% identical) to one of the polynucleotides of Figures 2-20 (SEQ ID NO:2-20), as well as fragments thereof.

In accordance with still another aspect of the present invention there is provided a polynucleotide which encodes for the same mature polypeptide as a human gene whose coding portion includes a polynucleotide which is at least 90% identical to (preferably at least 95% identical to and most preferably at least 97% or 100% identical) to one of the polynucleotides included in ATCC Deposit No. 97175 deposited June 2, 1995.

In accordance with yet another aspect of the present invention, there is provided a polynucleotide probe which hybridizes to mRNA (or the corresponding cDNA) which is transcribed from the coding portion of a human gene which coding portion includes a DNA sequence which is at least 90% identical to (preferably at least 95% identical to) and most preferably at least 97% or 100% identical) to one of the polynucleotide sequences of Figures 1-20 (SEQ ID NO:1-20).

The present invention further relates to a mature polypeptide encoded by a coding portion of a human gene which coding portion includes a DNA sequence which is at least 90%



identical to (preferably at least 95% identical to and more preferably 97% or 100% identical to) one of the polynucleotides of Figures 2-20 (SEQ ID NO:2-20), as well as analogues, derivatives and fragments of such polypeptides.

The present invention also relates to one of the mature polypeptides of Figure 1 (SEQ ID NO:1) and fragments, analogues and derivatives of such polypeptides.

The present invention further relates to the same mature polypeptide encoded by a human gene whose coding portion includes DNA which is at least 90% identical to (preferably at least 95% identical to and more preferably at least 97% or 100% identical to) one of the polynucleotides included in ATCC Deposit No. 97175 deposited June 2, 1995.

In accordance with an aspect of the present invention, there are provided isolated nucleic acids (polynucleotides) which encode for the mature polypeptides having the deduced amino acid sequence of Figure 1 (SEQ ID NO:1) or fragments, analogues or derivatives thereof.

The polynucleotides of the present invention may be in the form of RNA or in the form of DNA, which DNA includes cDNA, genomic DNA, and synthetic DNA. The DNA may be double-stranded or single-stranded, and if single stranded may be the coding strand or non-coding (anti-sense) strand. The coding sequence which encodes the mature polypeptide may include DNA identical to Figures 1-20 (SEQ ID NO:1-20) or that of the deposited clone or may be a different coding sequence which coding sequence, as a result of the redundancy or degeneracy of the genetic code, encodes the same mature polypeptide as the coding sequence of a gene which coding sequence includes the DNA of Figures 1-20 (SEQ ID NO:1-20) or the deposited cDNA.

The polynucleotide which encodes a mature polypeptide of the present invention may include, but is not limited to: only the coding sequence for the mature polypeptide; the coding sequence for the mature polypeptide and additional

coding sequence such as a leader or secretory sequence or a proprotein sequence; the coding sequence for the mature polypeptide (and optionally additional coding sequence) and non-coding sequence, such as introns or non-coding sequence 5' and/or 3' of the coding sequence for the mature polypeptide.

Thus, the term "polynucleotide encoding a polypeptide" encompasses a polynucleotide which includes only coding sequence for the polypeptide as well as a polynucleotide which includes additional coding and/or non-coding sequence.

The present invention further relates to variants of the hereinabove described polynucleotides which encode fragments, analogs and derivatives of a mature polypeptide of the present invention. The variant of the polynucleotide may be a naturally occurring allelic variant of the polynucleotide or a non-naturally occurring variant of the polynucleotide.

Thus, the present invention includes polynucleotides encoding the same mature polypeptide as hereinabove described as well as variants of such polynucleotides which variants encode a fragment, derivative or analog of a polypeptide of the invention. Such nucleotide variants include deletion variants, substitution variants and addition or insertion variants.

The polynucleotides of the invention may have a coding sequence which is a naturally occurring allelic variant of the human gene whose coding sequence includes DNA as shown in Figures 1-20 (SEQ ID NO:1-20) or of the coding sequence of the DNA in the deposited clone. As known in the art, an allelic variant is an alternate form of a polynucleotide sequence which may have a substitution, deletion or addition of one or more nucleotides, which does not substantially alter the function of the encoded polypeptide.

The present invention also includes polynucleotides, wherein the coding sequence for the mature polypeptide may be fused in the same reading frame to a polynucleotide sequence

which aids in expression and secretion of a polypeptide from a host cell, for example, a leader sequence which functions as a secretory sequence for controlling transport of a polypeptide from the cell. The polypeptide having a leader sequence is a preprotein and may have the leader sequence cleaved by the host cell to form the mature form of the polypeptide. The polynucleotides may also encode a proprotein which is the mature protein plus additional 5' amino acid residues. A mature protein having a prosequence is a preprotein and is an inactive form of the protein. Once the prosequence is cleaved an active mature protein remains.

Thus, for example, the polynucleotide of the present invention may encode a mature protein, or a protein having a prosequence or a protein having both a prosequence and a presequence (leader sequence).

The polynucleotides of the present invention may also have the coding sequence fused in frame to a marker sequence which allows for purification of the polypeptide of the present invention. The marker sequence may be a hexahistidine tag supplied by a pQE-9 vector to provide for purification of the mature polypeptide fused to the marker in the case of a bacterial host, or, for example, the marker sequence may be a hemagglutinin (HA) tag when a mammalian host, e.g. COS-7 cells, is used. The HA tag corresponds to an epitope derived from the influenza hemagglutinin protein (Wilson, I., et al., Cell, 37:767 (1984)).

The present invention further relates to polynucleotides which hybridize to the hereinabove-described polynucleotides if there is at least 70%, preferably at least 90%, and more preferably at least 95% identity between the sequences. The present invention particularly relates to polynucleotides which hybridize under stringent conditions to the hereinabove-described polynucleotides. As herein used, the term "stringent conditions" means hybridization will occur only if there is at least 95% and preferably at least

97% identity between the sequences. The polynucleotides which hybridize to the hereinabove described polynucleotides in a preferred embodiment encode polypeptides which retain substantially the same biological function or activity as the mature polypeptide of the present invention encoded by a coding sequence which includes the DNA of Figures 1-20 (SEQ ID NO:1-20) or the deposited cDNA(s).

Alternatively, the polynucleotide may have at least 10 or 20 bases, preferably at least 30 bases, and more preferably at least 50 bases which hybridize to a polynucleotide of the present invention and which has an identity thereto, as hereinabove described, and which may or may not retain activity. For example, such polynucleotides may be employed as probes for polynucleotides, for example, for recovery of the polynucleotide or as a diagnostic probe or as a PCR primer.

Thus, the present invention is directed to polynucleotides having at least a 70% identity, preferably at least 90% and more preferably at least 95% identity to a polynucleotide which encodes the mature polypeptide encoded by a human gene which includes the DNA of one of Figures 1-20 (SEQ ID NO:1-20) as well as fragments thereof, which fragments have at least 30 bases and preferably at least 50 bases and to polypeptides encoded by such polynucleotides.

The partial sequences are specific tags for messenger RNA molecules. The complete sequence of that messenger RNA, in the form of cDNA, can be determined using the partial sequence as a probe to identify a cDNA clone corresponding to a full-length transcript, followed by sequencing of that clone. The partial cDNA clone can also be used as a probe to identify a genomic clone or clones that contain the complete gene including regulatory and promoter regions, exons, and introns.

The partial sequences of Figures 2-20 (SEQ ID NO:2-20) may be used to identify the corresponding full length gene

from which they were derived. The partial sequences can be nick-translated or end-labelled with  $^{32}\text{P}$  using polynucleotide kinase using labelling methods known to those with skill in the art (Basic Methods in Molecular Biology, L.G. Davis, M.D. Dibner, and J.F. Battey, ed., Elsevier Press, NY, 1986). A lambda library prepared from human breast tissue can be directly screened with the labelled sequences of interest or the library can be converted en masse to pBluescript (Stratagene Cloning Systems, La Jolla, CA 92037) to facilitate bacterial breast screening. Regarding pBluescript, see Sambrook et al., Molecular Cloning-A Laboratory Manual, Cold Spring Harbor Laboratory Press (1989), pg. 1.20. Both methods are well known in the art. Briefly, filters with bacterial colonies containing the library in pBluescript or bacterial lawns containing lambda plaques are denatured and the DNA is fixed to the filters. The filters are hybridized with the labelled probe using hybridization conditions described by Davis et al., supra. The partial sequences, cloned into lambda or pBluescript, can be used as positive controls to assess background binding and to adjust the hybridization and washing stringencies necessary for accurate clone identification. The resulting autoradiograms are compared to duplicate plates of colonies or plaques; each exposed spot corresponds to a positive breast or plaque. The colonies or plaques are selected, expanded and the DNA is isolated from the colonies for further analysis and sequencing.

Positive cDNA clones are analyzed to determine the amount of additional sequence they contain using PCR with one primer from the partial sequence and the other primer from the vector. Clones with a larger vector-insert PCR product than the original partial sequence are analyzed by restriction digestion and DNA sequencing to determine whether they contain an insert of the same size or similar as the mRNA size determined from Northern blot Analysis.

Once one or more overlapping cDNA clones are identified, the complete sequence of the clones can be determined. The preferred method is to use exonuclease III digestion (McCombie, W.R, Kirkness, E., Fleming, J.T., Kerlavage, A.R., Iovannisci, D.M., and Martin-Gallardo, R., Methods, 3:33-40, 1991). A series of deletion clones are generated, each of which is sequenced. The resulting overlapping sequences are assembled into a single contiguous sequence of high redundancy (usually three to five overlapping sequences at each nucleotide position), resulting in a highly accurate final sequence.

The DNA sequences (as well as the corresponding RNA sequences) also include sequences which are or contain a DNA sequence identical to one contained in and isolatable from ATCC Deposit No. 97175, deposited June 2, 1995, and fragments or portions of the isolated DNA sequences (and corresponding RNA sequences), as well as DNA (RNA) sequences encoding the same polypeptide.

The deposit(s) referred to herein will be maintained under the terms of the Budapest Treaty on the International Recognition of the Deposit of Micro-organisms for purposes of Patent Procedure. These deposits are provided merely as convenience to those of skill in the art and are not an admission that a deposit is required under 35 U.S.C. §112. The sequence of the polynucleotides contained in the deposited materials, as well as the amino acid sequence of the polypeptides encoded thereby, are incorporated herein by reference and are controlling in the event of any conflict with any description of sequences herein. A license may be required to make, use or sell the deposited materials, and no such license is hereby granted.

The present invention further relates to polynucleotides which have at least 10 bases, preferably at least 20 bases, and may have 30 or more bases, which polynucleotides are hybridizable to and have at least a 70% identity to RNA (and

DNA which corresponds to such RNA) transcribed from a human gene whose coding portion includes DNA as hereinabove described.

Thus, the polynucleotide sequences which hybridize as described above may be used to hybridize to and detect the expression of the human genes to which they correspond for use in diagnostic assays as hereinafter described.

In accordance with still another aspect of the present invention there are provided diagnostic assays for detecting micrometastases of breast cancer in a host. While applicant does not wish to limit the reasoning of the present invention to any specific scientific theory, it is believed that the presence of active transcription of a breast specific gene of the present invention in cells of the host, other than those derived from the breast, is indicative of breast cancer metastases. This is true because, while the breast specific genes are found in all cells of the body, their transcription to mRNA, cDNA and expression products is primarily limited to the breast in non-diseased individuals. However, if breast cancer is present, breast cancer cells migrate from the cancer to other cells, such that these other cells are now actively transcribing and expressing a breast specific gene at a greater level than is normally found in non-diseased individuals, i.e., transcription is higher than found in non-breast tissues in healthy individuals. It is the detection of this enhanced transcription or enhanced protein expression in cells, other than those derived from the breast, which is indicative of metastases of breast cancer.

In one example of such a diagnostic assay, an RNA sequence in a sample derived from a tissue other than the breast is detected by hybridization to a probe. The sample contains a nucleic acid or a mixture of nucleic acids, at least one of which is suspected of containing a human breast specific gene or fragment thereof of the present invention which is transcribed and expressed in such tissue. Thus, for

example, in a form of an assay for determining the presence of a specific RNA in cells, initially RNA is isolated from the cells.

A sample may be obtained from cells derived from tissue other than from the breast including but not limited to blood, urine, saliva, tissue biopsy and autopsy material. The use of such methods for detecting enhanced transcription to mRNA from a human breast specific gene of the present invention or fragment thereof in a sample obtained from cells derived from other than the breast is well within the scope of those skilled in the art from the teachings herein.

The isolation of mRNA comprises isolating total cellular RNA by disrupting a cell and performing differential centrifugation. Once the total RNA is isolated, mRNA is isolated by making use of the adenine nucleotide residues known to those skilled in the art as a poly(A) tail found on virtually every eukaryotic mRNA molecule at the 3' end thereof. Oligonucleotides composed of only deoxythymidine [oligo(dT)] are linked to cellulose and the oligo(dT)-cellulose packed into small columns. When a preparation of total cellular RNA is passed through such a column, the mRNA molecules bind to the oligo(dT) by the poly(A) tails while the rest of the RNA flows through the column. The bound mRNAs are then eluted from the column and collected.

One example of detecting isolated mRNA transcribed from a breast specific gene of the present invention comprises screening the collected mRNAs with the gene specific oligonucleotide probes, as hereinabove described.

It is also appreciated that such probes can be and are preferably labeled with an analytically detectable reagent to facilitate identification of the probe. Useful reagents include but are not limited to radioactivity, fluorescent dyes or enzymes capable of catalyzing the formation of a detectable product.



An example of detecting a polynucleotide complementary to the mRNA sequence (cDNA) utilizes the polymerase chain reaction (PCR) in conjunction with reverse transcriptase. PCR is a very powerful method for the specific amplification of DNA or RNA stretches (Saiki et al., Nature, 234:163-166 (1986)). One application of this technology is in nucleic acid probe technology to bring up nucleic acid sequences present in low copy numbers to a detectable level. Numerous diagnostic and scientific applications of this method have been described by H.A. Erlich (ed.) in PCR Technology-Principles and Applications for DNA Amplification, Stockton Press, USA, 1989, and by M.A. Inis (ed.) in PCR Protocols, Academic Press, San Diego, USA, 1990.

RT-PCR is a combination of PCR with the reverse transcriptase enzyme. Reverse transcriptase is an enzyme which produces cDNA molecules from corresponding mRNA molecules. This is important since PCR amplifies nucleic acid molecules, particularly DNA, and this DNA may be produced from the mRNA isolated from a sample derived from the host.

A specific example of an RT-PCR diagnostic assay involves removing a sample from a tissue of a host. Such a sample will be from a tissue, other than the breast, for example, blood. Therefore, an example of such a diagnostic assay comprises whole blood gradient isolation of nucleated cells, total RNA extraction, RT-PCR of total RNA and agarose gel electrophoresis of PCR products. The PCR products comprise cDNA complementary to RNA transcribed from one or more breast specific genes of the present invention or fragments thereof. More particularly, a blood sample is obtained and the whole blood is combined with an equal volume of phosphate buffered saline, centrifuged and the lymphocyte and granulocyte layer is carefully aspirated and rediluted in phosphate buffered saline and centrifuged again. The supernate is discarded and the pellet containing nucleated

cells is used for RNA extraction using the RNazole B method as described by the manufacturer (Tel-Test Inc., Friendswood, TX).

Oligonucleotide primers and probes are prepared with high specificity to the DNA sequences of the present invention. The probes are at least 10 base pairs in length, preferably at least 30 base pairs in length and most preferably at least 50 base pairs in length or more. The reverse transcriptase reaction and PCR amplification are performed sequentially without interruption. Taq polymerase is used during PCR and the PCR products are concentrated and the entire sample is run on a Tris-borate-EDTA agarose gel containing ethidium bromide.

In accordance with another aspect of the present invention, there is provided a method of diagnosing a disorder of the breast, for example breast cancer, by determining altered levels of the breast specific polypeptides of the present invention in a biological sample, derived from tissue other than from the breast. Elevated levels of the breast specific polypeptides of the present invention, indicates active transcription and expression of the corresponding breast specific gene product. Assays used to detect levels of a breast specific gene polypeptide in a sample derived from a host are well-known to those skilled in the art and include radioimmunoassays, competitive-binding assays, Western blot analysis, ELISA assays and "sandwich" assays. A biological sample may include, but is not limited to, tissue extracts, cell samples or biological fluids, however, in accordance with the present invention, a biological sample specifically does not include tissue or cells of the breast.

An ELISA assay (Coligan, et al., Current Protocols in Immunology, 1(2), Chapter 6, 1991) initially comprises preparing an antibody specific to a breast specific polypeptide of the present invention, preferably a monoclonal

antibody. In addition, a reporter antibody is prepared against the monoclonal antibody. To the reporter antibody is attached a detectable reagent such as radioactivity, fluorescence or, in this example, a horseradish peroxidase enzyme. A sample is removed from a host and incubated on a solid support, e.g., a polystyrene dish, that binds the proteins in the sample. Any free protein binding sites on the dish are then covered by incubating with a non-specific protein, such as BSA. Next, the monoclonal antibody is incubated in the dish during which time the monoclonal antibodies attach to the breast specific polypeptide attached to the polystyrene dish. All unbound monoclonal antibody is washed out with buffer. The reporter antibody linked to horseradish peroxidase is now placed in the dish resulting in binding of the reporter antibody to any monoclonal antibody bound to the breast specific gene polypeptide. Unattached reporter antibody is then washed out. Peroxidase substrates are then added to the dish and the amount of color developed in a given time period is a measurement of the amount of the breast specific polypeptide present in a given volume of patient sample when compared against a standard curve.

A competition assay may be employed where antibodies specific to a breast specific polypeptide are attached to a solid support. The breast specific polypeptide is then labeled and the labeled polypeptide a sample derived from the host are passed over the solid support and the amount of label detected, for example, by liquid scintillation chromatography, can be correlated to a quantity of the breast specific polypeptide in the sample.

A "sandwich" assay is similar to an ELISA assay. In a "sandwich" assay, breast specific polypeptides are passed over a solid support and bind to antibody attached to the solid support. A second antibody is then bound to the breast specific polypeptide. A third antibody which is labeled and is specific to the second antibody, is then passed over the

solid support and binds to the second antibody and an amount can then be quantified.

In alternative methods, labeled antibodies to a breast specific polypeptide are used. In a one-step assay, the target molecule, if it is present, is immobilized and incubated with a labeled antibody. The labeled antibody binds to the immobilized target molecule. After washing to remove the unbound molecules, the sample is assayed for the presence of the label. In a two-step assay, immobilized target molecule is incubated with an unlabeled antibody. The target molecule-labeled antibody complex, if present, is then bound to a second, labeled antibody that is specific for the unlabeled antibody. The sample is washed and assayed for the presence of the label.

Such antibodies specific to breast specific gene proteins, for example, anti-idiotypic antibodies, can be used to detect breast cancer cells by being labeled and described above and binding tightly to the breast cancer cells, and, therefore, detect their presence.

The antibodies may also be used to target breast cancer cells, for example, in a method of homing interaction agents which, when contacting breast cancer cells, destroy them. This is true since the antibodies are specific for breast specific genes which are primarily expressed in breast cancer, and a linking of the interaction agent to the antibody would cause the interaction agent to be carried directly to the breast.

Antibodies of this type may also be used to do *in vivo* imaging, for example, by labeling the antibodies to facilitate scanning of the breast. One method for imaging comprises contacting any cancer cells of the breast to be imaged with an anti-breast specific gene protein antibody labeled with a detectable marker. The method is performed under conditions such that the labeled antibody binds to the breast specific gene proteins. In a specific example, the

antibodies interact with the breast, for example, breast cancer cells, and fluoresce upon such contact such that imaging and visibility of the breast is enhanced to allow a determination of the diseased or non-diseased state of the breast.

The choice of marker used to label the antibodies will vary depending upon the application. However, the choice of marker is readily determinable to one skilled in the art. These labeled antibodies may be used in immunoassays as well as in histological applications to detect the presence of the proteins. The labeled antibodies may be polyclonal or monoclonal.

The presence of active transcription, which is greater than that normally found, of the breast specific genes in cells other than from the breast, by the presence of an altered level of mRNA, cDNA or expression products is an important indication of the presence of a breast cancer which has metastasized, since breast cancer cells are migrating from the breast into the general circulation. Accordingly, this phenomenon may have important clinical implications since the method of treating a localized, as opposed to a metastasized, tumor is entirely different.

Of the 20 breast specific genes disclosed, only breast specific gene 1 is a full-length gene. Breast specific gene 1 is 79% identical and 83% similar to human Alzheimer disease amyloid gene. Breast specific gene 2 is 30% identical and 48% similar to human hydroxyindole-o-methyltransferase gene. Breast specific gene 3 is 58% identical and 62% similar to human 06-methylguanine-DNA methyltransferase gene. Breast specific gene 4 is 34% identical and 65% similar to the mouse p120 gene. Breast specific gene 5 is 78% identical and 89% similar to human p70 ribosomal S6 kinase alpha-II gene. Breast specific gene 6 is 77% identical and 79% similar to the human transcription factor NFATp gene.

As stated previously, the breast specific genes of the present invention are putative molecular markers in the diagnosis of breast cancer formation, and breast cancer metastases. As shown in the following Table 1, the presence of the breast specific genes when tested in normal breast, breast cancer, embryo and other cancer libraries, the breast specific genes of the present invention were found to be most prevalent in the breast cancer library, indicating that the genes of the present invention may be employed for detecting breast cancer, as discussed previously. The table also indicates a putative identification, based on homology, of BSG1 through BSG6 to known genes.

Table 1

Genes	Homolog Gene Name (Class)	Norm Br	Br Ca	Embryo	Other Cancers	Others
BSG1	AD Amyloid (3)	1	6			1
BSG2	Hydroxyindole-o-methyltransferase (2)		3	1	1	
BSG3	O-6-methylguanine-DNA methyltransferase (1)		3	1	1	
BSG4	P120 (3)		3		1	
BSG5	p70 ribosomal S6 kinase alpha-II (2)		3		1	
BSG6	Transcription factor NFATp (3)	2				
BSG7			2	1		
BSG8			4		3	1
BSG8			2			
BSG9			3			
BSG10			3			
BSG11			3			
BSG12		3	3			
BSG13			3			

BSG14			2			
BSG15		3				
BSG16		1	1	1		
BSG17			2		1	
BSG18		2				
BSG19		1	1			
BSG20			2			

The assays described above may also be used to test whether bone marrow preserved before chemotherapy is contaminated with micrometastases of a breast cancer cell. In the assay, blood cells from the bone marrow are isolated and treated as described above, this method allows one to determine whether preserved bone marrow is still suitable for transplantation after chemotherapy.

The present invention further relates to mature polypeptides, for example the BSG1 polypeptide, as well as fragments, analogs and derivatives of such polypeptide.

The terms "fragment," "derivative" and "analog" when referring to the polypeptides encoded by the genes of the invention means a polypeptide which retains essentially the same biological function or activity as such polypeptide. Thus, an analog includes a proprotein which can be activated by cleavage of the proprotein portion to produce an active mature polypeptide.

The polypeptides of the present invention may be recombinant polypeptides, natural polypeptides or synthetic polypeptides, preferably recombinant polypeptides.

The fragment, derivative or analog of the polypeptides encoded by the genes of the invention may be (i) one in which one or more of the amino acid residues are substituted with a conserved or non-conserved amino acid residue (preferably a conserved amino acid residue) and such substituted amino acid residue may or may not be one encoded by the genetic code, or (ii) one in which one or more of the amino acid

residues includes a substituent group, or (iii) one in which the polypeptide is fused with another compound, such as a compound to increase the half-life of the polypeptide (for example, polyethylene glycol), or (iv) one in which the additional amino acids are fused to the polypeptide, such as a leader or secretory sequence or a sequence which is employed for purification of the mature polypeptide or a proprotein sequence. Such fragments, derivatives and analogs are deemed to be within the scope of those skilled in the art from the teachings herein.

The polypeptides and polynucleotides of the present invention are preferably provided in an isolated form, and preferably are purified to homogeneity.

The term "isolated" means that the material is removed from its original environment (e.g., the natural environment if it is naturally occurring). For example, a naturally-occurring polynucleotide or polypeptide present in a living animal is not isolated, but the same polynucleotide or polypeptide, separated from some or all of the coexisting materials in the natural system, is isolated. Such polynucleotides could be part of a vector and/or such polynucleotides or polypeptides could be part of a composition, and still be isolated in that such vector or composition is not part of its natural environment.

The polypeptides of the present invention include the polypeptides of Figure 1 (SEQ ID NO:1) (in particular the mature polypeptides) as well as polypeptides which have at least 70% similarity (preferably at least a 70% identity) to the polypeptides of Figure 1 (SEQ ID NO:1) and more preferably at least a 90% similarity (more preferably at least a 90% identity) to the polypeptides of Figures 8 and 9 (SEQ ID NO:8 and 9) and still more preferably at least a 95% similarity (still more preferably at least 95% identity) to the polypeptides of Figure 1 (SEQ ID NO:1) and also include portions of such polypeptides with such portion of the



polypeptide generally containing at least 30 amino acids and more preferably at least 50 amino acids.

As known in the art "similarity" between two polypeptides is determined by comparing the amino acid sequence and its conserved amino acid substitutes of one polypeptide to the sequence of a second polypeptide.

Fragments or portions of the polypeptides of the present invention may be employed for producing the corresponding full-length polypeptide by peptide synthesis; therefore, the fragments may be employed as intermediates for producing the full-length polypeptides. Fragments or portions of the polynucleotides of the present invention may be used to synthesize full-length polynucleotides of the present invention.

The present invention also relates to vectors which include polynucleotides of the present invention, host cells which are genetically engineered with vectors of the invention and the production of polypeptides of the invention by recombinant techniques.

Host cells are genetically engineered (transduced or transformed or transfected) with the vectors of this invention which may be, for example, a cloning vector or an expression vector. The vector may be, for example, in the form of a plasmid, a viral particle, a phage, etc. The engineered host cells can be cultured in conventional nutrient media modified as appropriate for activating promoters, selecting transformants or amplifying the breast specific genes. The culture conditions, such as temperature, pH and the like, are those previously used with the host cell selected for expression, and will be apparent to those of ordinary skill in the art.

The polynucleotides of the present invention may be employed for producing polypeptides by recombinant techniques. Thus, for example, the polynucleotide may be included in any one of a variety of expression vectors for

expressing a polypeptide. Such vectors include chromosomal, nonchromosomal and synthetic DNA sequences, e.g., derivatives of SV40; bacterial plasmids; phage DNA; baculovirus; yeast plasmids; vectors derived from combinations of plasmids and phage DNA, viral DNA such as vaccinia, adenovirus, fowl pox virus, and pseudorabies. However, any other vector may be used as long as it is replicable and viable in the host.

The appropriate DNA sequence may be inserted into the vector by a variety of procedures. In general, the DNA sequence is inserted into an appropriate restriction endonuclease site(s) by procedures known in the art. Such procedures and others are deemed to be within the scope of those skilled in the art.

The DNA sequence in the expression vector is operatively linked to an appropriate expression control sequence(s) (promoter) to direct mRNA synthesis. As representative examples of such promoters, there may be mentioned: LTR or SV40 promoter, the E. coli lac or trp, the phage lambda P<sub>L</sub> promoter and other promoters known to control expression of genes in prokaryotic or eukaryotic cells or their viruses. The expression vector also contains a ribosome binding site for translation initiation and a transcription terminator. The vector may also include appropriate sequences for amplifying expression.

In addition, the expression vectors preferably contain one or more selectable marker genes to provide a phenotypic trait for selection of transformed host cells such as dihydrofolate reductase or neomycin resistance for eukaryotic cell culture, or such as tetracycline or ampicillin resistance in E. coli.

The vector containing the appropriate DNA sequence as hereinabove described, as well as an appropriate promoter or control sequence, may be employed to transform an appropriate host to permit the host to express the protein.

As representative examples of appropriate hosts, there may be mentioned: bacterial cells, such as E. coli, Streptomyces, Salmonella typhimurium; fungal cells, such as yeast; insect cells such as Drosophila S2 and Spodoptera Sf9; animal cells such as CHO, COS or Bowes melanoma; adenoviruses; plant cells, etc. The selection of an appropriate host is deemed to be within the scope of those skilled in the art from the teachings herein.

More particularly, the present invention also includes recombinant constructs comprising one or more of the sequences as broadly described above. The constructs comprise a vector, such as a plasmid or viral vector, into which a sequence of the invention has been inserted, in a forward or reverse orientation. In a preferred aspect of this embodiment, the construct further comprises regulatory sequences, including, for example, a promoter, operably linked to the sequence. Large numbers of suitable vectors and promoters are known to those of skill in the art, and are commercially available. The following vectors are provided by way of example. Bacterial: pQE70, pQE60, pQE-9 (Qiagen), pBS, pD10, phagescript, psiX174, pbluescript SK, pBSKS, pNH8A, pNH16a, pNH18A, pNH46A (Stratagene); ptrc99a, pKK223-3, pKK233-3, pDR540, pRIT5 (Pharmacia). Eukaryotic: pWLNEO, pSV2CAT, pOG44, pXT1, pSG (Stratagene) pSVK3, pBPV, pMSG, pSVL (Pharmacia). However, any other plasmid or vector may be used as long as they are replicable and viable in the host.

Promoter regions can be selected from any desired gene using CAT (chloramphenicol transferase) vectors or other vectors with selectable markers. Two appropriate vectors are pKK232-8 and pCM7. Particular named bacterial promoters include lacI, lacZ, T3, T7, gpt, lambda P<sub>R</sub>, P<sub>L</sub> and trp. Eukaryotic promoters include CMV immediate early, HSV thymidine kinase, early and late SV40, LTRs from retrovirus, and mouse metallothionein-I. Selection of the appropriate

vector and promoter is well within the level of ordinary skill in the art.

In a further embodiment, the present invention relates to host cells containing the above-described constructs. The host cell can be a higher eukaryotic cell, such as a mammalian cell, or a lower eukaryotic cell, such as a yeast cell, or the host cell can be a prokaryotic cell, such as a bacterial cell. Introduction of the construct into the host cell can be effected by calcium phosphate transfection, DEAE-Dextran mediated transfection, or electroporation (Davis, L., Dibner, M., Battey, I., Basic Methods in Molecular Biology, (1986)).

The constructs in host cells can be used in a conventional manner to produce the gene product encoded by the recombinant sequence. Alternatively, the polypeptides of the invention can be synthetically produced by conventional peptide synthesizers.

Proteins can be expressed in mammalian cells, yeast, bacteria, or other cells under the control of appropriate promoters. Cell-free translation systems can also be employed to produce such proteins using RNAs derived from the DNA constructs of the present invention. Appropriate cloning and expression vectors for use with prokaryotic and eukaryotic hosts are described by Sambrook, et al., Molecular Cloning: A Laboratory Manual, Second Edition, Cold Spring Harbor, N.Y., (1989), the disclosure of which is hereby incorporated by reference.

Transcription of the DNA encoding the polypeptides of the present invention by higher eukaryotes is increased by inserting an enhancer sequence into the vector. Enhancers are cis-acting elements of DNA, usually about from 10 to 300 bp that act on a promoter to increase its transcription. Examples including the SV40 enhancer on the late side of the replication origin bp 100 to 270, a cytomegalovirus early

promoter enhancer, the polyoma enhancer on the late side of the replication origin, and adenovirus enhancers.

Generally, recombinant expression vectors will include origins of replication and selectable markers permitting transformation of the host cell, e.g., the ampicillin resistance gene of E. coli and S. cerevisiae TRP1 gene, and a promoter derived from a highly-expressed gene to direct transcription of a downstream structural sequence. Such promoters can be derived from operons encoding glycolytic enzymes such as 3-phosphoglycerate kinase (PGK),  $\alpha$ -factor, acid phosphatase, or heat shock proteins, among others. The heterologous structural sequence is assembled in appropriate phase with translation initiation and termination sequences. Optionally, the heterologous sequence can encode a fusion protein including an N-terminal identification peptide imparting desired characteristics, e.g., stabilization or simplified purification of expressed recombinant product.

Useful expression vectors for bacterial use are constructed by inserting a structural DNA sequence encoding a desired protein together with suitable translation initiation and termination signals in operable reading frame with a functional promoter. The vector will comprise one or more phenotypic selectable markers and an origin of replication to ensure maintenance of the vector and to, if desirable, provide amplification within the host. Suitable prokaryotic hosts for transformation include E. coli, Bacillus subtilis, Salmonella typhimurium and various species within the genera Pseudomonas, Streptomyces, and Staphylococcus, although others may also be employed as a matter of choice.

As a representative but nonlimiting example, useful expression vectors for bacterial use can comprise a selectable marker and bacterial origin of replication derived from commercially available plasmids comprising genetic elements of the well known cloning vector pBR322 (ATCC

37017). Such commercial vectors include, for example, pKK223-3 (Pharmacia Fine Chemicals, Uppsala, Sweden) and GEM1 (Promega Biotec, Madison, WI, USA). These pBR322 "backbone" sections are combined with an appropriate promoter and the structural sequence to be expressed.

Following transformation of a suitable host strain and growth of the host strain to an appropriate cell density, the selected promoter is induced by appropriate means (e.g., temperature shift or chemical induction) and cells are cultured for an additional period.

Cells are typically harvested by centrifugation, disrupted by physical or chemical means, and the resulting crude extract retained for further purification.

Microbial cells employed in expression of proteins can be disrupted by any convenient method, including freeze-thaw cycling, sonication, mechanical disruption, or use of cell lysing agents, such methods are well known to those skilled in the art.

Various mammalian cell culture systems can also be employed to express recombinant protein. Examples of mammalian expression systems include the COS-7 lines of monkey kidney fibroblasts, described by Gluzman, *Cell*, 23:175 (1981), and other cell lines capable of expressing a compatible vector, for example, the C127, 3T3, CHO, HeLa and BHK cell lines. Mammalian expression vectors will comprise an origin of replication, a suitable promoter and enhancer, and also any necessary ribosome binding sites, polyadenylation site, splice donor and acceptor sites, transcriptional termination sequences, and 5' flanking nontranscribed sequences. DNA sequences derived from the SV40 splice, and polyadenylation sites may be used to provide the required nontranscribed genetic elements.

The breast specific gene polypeptides can be recovered and purified from recombinant cell cultures by methods including ammonium sulfate or ethanol precipitation, acid

extraction, anion or cation exchange chromatography, phosphocellulose chromatography, hydrophobic interaction chromatography, affinity chromatography, hydroxylapatite chromatography and lectin chromatography. Protein refolding steps can be used, as necessary, in completing configuration of the mature protein. Finally, high performance liquid chromatography (HPLC) can be employed for final purification steps.

The polynucleotides of the present invention may have the coding sequence fused in frame to a marker sequence which allows for purification of the polypeptide of the present invention. An example of a marker sequence is a hexahistidine tag which may be supplied by a vector, preferably a pQE-9 vector, which provides for purification of the polypeptide fused to the marker in the case of a bacterial host, or, for example, the marker sequence may be a hemagglutinin (HA) tag when a mammalian host, e.g. COS-7 cells, is used. The HA tag corresponds to an epitope derived from the influenza hemagglutinin protein (Wilson, I., et al., Cell, 37:767 (1984)).

The polypeptides of the present invention may be a naturally purified product, or a product of chemical synthetic procedures, or produced by recombinant techniques from a prokaryotic or eukaryotic host (for example, by bacterial, yeast, higher plant, insect and mammalian cells in culture). Depending upon the host employed in a recombinant production procedure, the polypeptides of the present invention may be glycosylated or may be non-glycosylated. Polypeptides of the invention may also include an initial methionine amino acid residue.

BSG1, and other breast specific genes, and the protein product thereof may be employed for early detection of breast cancer since they are over-expressed in the breast cancer state.

In accordance with another aspect of the present invention there are provided assays which may be used to screen for therapeutics to inhibit the action of the breast specific genes or breast specific proteins of the present invention. The present invention discloses methods for selecting a therapeutic which forms a complex with breast specific gene proteins with sufficient affinity to prevent their biological action. The methods include various assays, including competitive assays where the proteins are immobilized to a support, and are contacted with a natural substrate and a labeled therapeutic either simultaneously or in either consecutive order, and determining whether the therapeutic effectively competes with the natural substrate in a manner sufficient to prevent binding of the protein to its substrate.

In another embodiment, the substrate is immobilized to a support, and is contacted with both a labeled breast specific polypeptide and a therapeutic (or unlabeled proteins and a labeled therapeutic), and it is determined whether the amount of the breast specific polypeptide bound to the substrate is reduced in comparison to the assay without the therapeutic added. The breast specific polypeptide may be labeled with antibodies.

Potential therapeutic compounds include antibodies and anti-idiotypic antibodies as described above, or in some cases, an oligonucleotide, which binds to the polypeptide.

Another example is an antisense construct prepared using antisense technology, which is directed to a breast specific polynucleotide to prevent transcription. Antisense technology can be used to control gene expression through triple-helix formation or antisense DNA or RNA, both of which methods are based on binding of a polynucleotide to DNA or RNA. For example, the 5' coding portion of the polynucleotide sequence, which encodes for the mature polypeptides of the present invention, is used to design an



antisense RNA oligonucleotide of from about 10 to 40 base pairs in length. A DNA oligonucleotide is designed to be complementary to a region of the gene involved in transcription (triple helix -see Lee et al., Nucl. Acids Res., 6:3073 (1979); Cooney et al, Science, 241:456 (1988); and Dervan et al., Science, 251: 1360 (1991)), thereby preventing transcription and the production of a breast specific polynucleotide. The antisense RNA oligonucleotide hybridizes to the mRNA in vivo and blocks translation of the mRNA molecule into the breast specific genes polypeptide (antisense - Okano, J. Neurochem., 56:560 (1991); Oligodeoxynucleotides as Antisense Inhibitors of Gene Expression, CRC Press, Boca Raton, FL (1988)). The oligonucleotides described above can also be delivered to cells such that the antisense RNA or DNA may be expressed in vivo to inhibit production of the breast specific polypeptides.

Another example is a small molecule which binds to and occupies the active site of the breast specific polypeptide thereby making the active site inaccessible to substrate such that normal biological activity is prevented. Examples of small molecules include but are not limited to small peptides or peptide-like molecules.

These compounds may be employed to treat breast cancer, since they interact with the function of breast specific polypeptides in a manner sufficient to inhibit natural function which is necessary for the viability of breast cancer cells. This is true since the BSGs and their protein products are primarily expressed in breast cancer tissues and are, therefore, suspected of being critical to the formation of this state.

The compounds may be employed in a composition with a pharmaceutically acceptable carrier, e.g., as hereinafter described.

The compounds of the present invention may be employed in combination with a suitable pharmaceutical carrier. Such compositions comprise a therapeutically effective amount of the polypeptide, and a pharmaceutically acceptable carrier or excipient. Such a carrier includes but is not limited to saline, buffered saline, dextrose, water, glycerol, ethanol, and combinations thereof. The formulation should suit the mode of administration.

The invention also provides a pharmaceutical pack or kit comprising one or more containers filled with one or more of the ingredients of the pharmaceutical compositions of the invention. Associated with such container(s) can be a notice in the form prescribed by a governmental agency regulating the manufacture, use or sale of pharmaceuticals or biological products, which notice reflects approval by the agency of manufacture, use or sale for human administration. In addition, the pharmaceutical compositions may be employed in conjunction with other therapeutic compounds.

The pharmaceutical compositions may be administered in a convenient manner such as by the oral, topical, intravenous, intraperitoneal, intramuscular, subcutaneous, intranasal, intra-anal or intradermal routes. The pharmaceutical compositions are administered in an amount which is effective for treating and/or prophylaxis of the specific indication. In general, they are administered in an amount of at least about 10  $\mu\text{g/kg}$  body weight and in most cases they will be administered in an amount not in excess of about 8 mg/Kg body weight per day. In most cases, the dosage is from about 10  $\mu\text{g/kg}$  to about 1 mg/kg body weight daily, taking into account the routes of administration, symptoms, etc.

The breast specific genes and compounds which are polypeptides may also be employed in accordance with the present invention by expression of such polypeptides *in vivo*, which is often referred to as "gene therapy."

Thus, for example, cells from a patient may be engineered with a polynucleotide (DNA or RNA) encoding a polypeptide *ex vivo*, with the engineered cells then being provided to a patient to be treated with the polypeptide. Such methods are well-known in the art. For example, cells may be engineered by procedures known in the art by use of a retroviral particle containing RNA encoding a polypeptide of the present invention.

Similarly, cells may be engineered *in vivo* for expression of a polypeptide *in vivo* by, for example, procedures known in the art. As known in the art, a producer cell for producing a retroviral particle containing RNA encoding a polypeptide of the present invention may be administered to a patient for engineering cells *in vivo* and expression of the polypeptide *in vivo*. These and other methods for administering a polypeptide of the present invention by such method should be apparent to those skilled in the art from the teachings of the present invention. For example, the expression vehicle for engineering cells may be other than a retrovirus, for example, an adenovirus which may be used to engineer cells *in vivo* after combination with a suitable delivery vehicle.

Retroviruses from which the retroviral plasmid vectors hereinabove mentioned may be derived include, but are not limited to, Moloney Murine Leukemia Virus, spleen necrosis virus, retroviruses such as Rous Sarcoma Virus, Harvey Sarcoma Virus, avian leukosis virus, gibbon ape leukemia virus, human immunodeficiency virus, adenovirus, Myeloproliferative Sarcoma Virus, and mammary tumor virus. In one embodiment, the retroviral plasmid vector is derived from Moloney Murine Leukemia Virus.

The vector includes one or more promoters. Suitable promoters which may be employed include, but are not limited to, the retroviral LTR; the SV40 promoter; and the human cytomegalovirus (CMV) promoter described in Miller, et al.,

Biotechniques, Vol. 7, No. 9, 980-990 (1989), or any other promoter (e.g., cellular promoters such as eukaryotic cellular promoters including, but not limited to, the histone, pol III, and  $\beta$ -actin promoters). Other viral promoters which may be employed include, but are not limited to, adenovirus promoters, thymidine kinase (TK) promoters, and B19 parvovirus promoters. The selection of a suitable promoter will be apparent to those skilled in the art from the teachings contained herein.

The nucleic acid sequence encoding the polypeptide of the present invention is under the control of a suitable promoter. Suitable promoters which may be employed include, but are not limited to, adenoviral promoters, such as the adenoviral major late promoter; or heterologous promoters, such as the cytomegalovirus (CMV) promoter; the respiratory syncytial virus (RSV) promoter; inducible promoters, such as the MMT promoter, the metallothionein promoter; heat shock promoters; the albumin promoter; the ApoAI promoter; human globin promoters; viral thymidine kinase promoters, such as the Herpes Simplex thymidine kinase promoter; retroviral LTRs (including the modified retroviral LTRs hereinabove described); the  $\beta$ -actin promoter; and human growth hormone promoters. The promoter also may be the native promoter which controls the genes encoding the polypeptides.

The retroviral plasmid vector is employed to transduce packaging cell lines to form producer cell lines. Examples of packaging cells which may be transfected include, but are not limited to, the PE501, PA317,  $\psi$ -2,  $\psi$ -AM, PA12, T19-14X, VT-19-17-H2,  $\psi$ CRE,  $\psi$ CRIP, GP+E-86, GP+envAm12, and DAN cell lines as described in Miller, Human Gene Therapy, Vol. 1, pgs. 5-14 (1990), which is incorporated herein by reference in its entirety. The vector may transduce the packaging cells through any means known in the art. Such means include, but are not limited to, electroporation, the use of liposomes, and CaPO<sub>4</sub> precipitation. In one alternative, the

retroviral plasmid vector may be encapsulated into a liposome, or coupled to a lipid, and then administered to a host.

The producer cell line generates infectious retroviral vector particles which include the nucleic acid sequence(s) encoding the polypeptides. Such retroviral vector particles then may be employed, to transduce eukaryotic cells, either *in vitro* or *in vivo*. The transduced eukaryotic cells will express the nucleic acid sequence(s) encoding the polypeptide. Eukaryotic cells which may be transduced include, but are not limited to, embryonic stem cells, embryonic carcinoma cells, as well as hematopoietic stem cells, hepatocytes, fibroblasts, myoblasts, keratinocytes, endothelial cells, and bronchial epithelial cells.

This invention is also related to the use of a breast specific genes of the present invention as a diagnostic. For example, some diseases result from inherited defective genes. The breast specific genes, CSG7 and CSG10, for example, have been found to have a reduced expression in breast cancer cells as compared to that in normal cells. Further, the remaining breast specific genes of the present invention are overexpressed in breast cancer. Accordingly, a mutation in these genes allows a detection of breast disorders, for example, breast cancer. A mutation in a breast specific gene of the present invention at the DNA level may be detected by a variety of techniques. Nucleic acids used for diagnosis (genomic DNA, mRNA, etc.) may be obtained from a patient's cells, other than from the breast, such as from blood, urine, saliva, tissue biopsy and autopsy material. The genomic DNA may be used directly for detection or may be amplified enzymatically by using PCR (Saiki, et al., Nature, 324:163-166 (1986)) prior to analysis. RNA or cDNA may also be used for the same purpose. As an example, PCR primers complementary to the nucleic acid of the instant invention can be used to identify and analyze mutations in a breast

specific polynucleotide of the present invention. For example, deletions and insertions can be detected by a change in size of the amplified product in comparison to the normal genotype. Point mutations can be identified by hybridizing amplified DNA to radiolabelled breast specific RNA or, alternatively, radiolabelled antisense DNA sequences.

Another well-established method for screening for mutations in particular segments of DNA after PCR amplification is single-strand conformation polymorphism (SSCP) analysis. PCR products are prepared for SSCP by ten cycles of reamplification to incorporate  $^{32}\text{P}$ -dCTP, digested with an appropriate restriction enzyme to generate 200-300 bp fragments, and denatured by heating to 85°C for 5 min. and then plunged into ice. Electrophoresis is then carried out in a nondenaturing gel (5% glycerol, 5% acrylamide) (Glavac, D. and Dean, M., Human Mutation, 2:404-414 (1993)).

Sequence differences between the reference gene and "mutants" may be revealed by the direct DNA sequencing method. In addition, cloned DNA segments may be used as probes to detect specific DNA segments. The sensitivity of this method is greatly enhanced when combined with PCR. For example, a sequencing primer is used with double-stranded PCR product or a single-stranded template molecule generated by a modified PCR. The sequence determination is performed by conventional procedures with radiolabeled nucleotides or by automatic sequencing procedures with fluorescent-tags.

Genetic testing based on DNA sequence differences may be achieved by detection of alteration in electrophoretic mobility of DNA fragments and gels with or without denaturing agents. Small sequence deletions and insertions can be visualized by high-resolution gel electrophoresis. DNA fragments of different sequences may be distinguished on denaturing formamide gradient gels in which the mobilities of different DNA fragments are retarded in the gel at different positions according to their specific melting or partial

melting temperatures (see, e.g., Myers, et al., Science, 230:1242 (1985)). In addition, sequence alterations, in particular small deletions, may be detected as changes in the migration pattern of DNA.

Sequence changes at specific locations may also be revealed by nuclease protection assays, such as Rnase and S1 protection or the chemical cleavage method (e.g., Cotton, et al., PNAS, USA, 85:4397-4401 (1985)).

Thus, the detection of the specific DNA sequence may be achieved by methods such as hybridization, RNase protection, chemical cleavage, direct DNA sequencing, or the use of restriction enzymes (e.g., Restriction Fragment Length Polymorphisms (RFLP)) and Southern blotting.

The sequences of the present invention are also valuable for chromosome identification. The sequence is specifically targeted to and can hybridize with a particular location on an individual human chromosome. Moreover, there is a current need for identifying particular sites on the chromosome. Few chromosome marking reagents based on actual sequence data (repeat polymorphisms) are presently available for marking chromosomal location. The mapping of DNAs to chromosomes according to the present invention is an important first step in correlating those sequences with genes associated with disease.

Briefly, sequences can be mapped to chromosomes by preparing PCR primers (preferably 15-25 bp) from the cDNA. Computer analysis of the 3' untranslated region is used to rapidly select primers that do not span more than one exon in the genomic DNA, thus complicating the amplification process. These primers are then used for PCR screening of somatic cell hybrids containing individual human chromosomes. Only those hybrids containing the human gene corresponding to the primer will yield an amplified fragment.

PCR mapping of somatic cell hybrids is a rapid procedure for assigning a particular DNA to a particular chromosome.

Using the present invention with the same oligonucleotide primers, sublocalization can be achieved with panels of fragments from specific chromosomes or pools of large genomic clones in an analogous manner. Other mapping strategies that can similarly be used to map to its chromosome include *in situ* hybridization, prescreening with labeled flow-sorted chromosomes and preselection by hybridization to construct chromosome specific-cDNA libraries.

Fluorescence *in situ* hybridization (FISH) of a cDNA clone to a metaphase chromosomal spread can be used to provide a precise chromosomal location in one step. This technique can be used with cDNA as short as 50 or 60 bases. For a review of this technique, see Verma et al., Human Chromosomes: a Manual of Basic Techniques, Pergamon Press, New York (1988).

Once a sequence has been mapped to a precise chromosomal location, the physical position of the sequence on the chromosome can be correlated with genetic map data. Such data are found, for example, in V. McKusick, Mendelian Inheritance in Man (available on line through Johns Hopkins University Welch Medical Library). The relationship between genes and diseases that have been mapped to the same chromosomal region are then identified through linkage analysis (coinheritance of physically adjacent genes).

Next, it is necessary to determine the differences in the cDNA or genomic sequence between affected and unaffected individuals. If a mutation is observed in some or all of the affected individuals but not in any normal individuals, then the mutation is likely to be the causative agent of the disease.

With current resolution of physical mapping and genetic mapping techniques, a cDNA precisely localized to a chromosomal region associated with the disease, could be one of between 50 and 500 potential causative genes. (This



assumes 1 megabase mapping resolution and one gene per 20 kb).

The polypeptides, their fragments or other derivatives, or analogs thereof, or cells expressing them can be used as an immunogen to produce antibodies thereto. These antibodies can be, for example, polyclonal or monoclonal antibodies. The present invention also includes chimeric, single chain, and humanized antibodies, as well as Fab fragments, or the product of an Fab expression library. Various procedures known in the art may be used for the production of such antibodies and fragments.

Antibodies generated against the polypeptides corresponding to a sequence of the present invention can be obtained by direct injection of the polypeptides into an animal or by administering the polypeptides to an animal, preferably a nonhuman. The antibody so obtained will then bind the polypeptides itself. In this manner, even a sequence encoding only a fragment of the polypeptides can be used to generate antibodies binding the whole native polypeptides. Such antibodies can then be used to isolate the polypeptide from tissue expressing that polypeptide.

For preparation of monoclonal antibodies, any technique which provides antibodies produced by continuous cell line cultures can be used. Examples include the hybridoma technique (Kohler and Milstein, 1975, *Nature*, 256:495-497), the trioma technique, the human B-cell hybridoma technique (Kozbor et al., 1983, *Immunology Today* 4:72), and the EBV-hybridoma technique to produce human monoclonal antibodies (Cole, et al., 1985, in *Monoclonal Antibodies and Cancer Therapy*, Alan R. Liss, Inc., pp. 77-96).

Techniques described for the production of single chain antibodies (U.S. Patent 4,946,778) can be adapted to produce single chain antibodies to immunogenic polypeptide products of this invention. Transgenic mice may also be used to generate antibodies.

The antibodies may also be employed to target breast cancer cells, for example, in a method of homing interaction agents which, when contacting breast cancer cells, destroy them. This is true since the antibodies are specific for the breast specific polypeptides of the present invention. A linking of the interaction agent to the antibody would cause the interaction agent to be carried directly to the breast.

Antibodies of this type may also be used to do *in vivo* imaging, for example, by labeling the antibodies to facilitate scanning of the pelvic area and the breast. One method for imaging comprises contacting any cancer cells of the breast to be imaged with an anti-breast specific protein-antibody labeled with a detectable marker. The method is performed under conditions such that the labeled antibody binds to the breast specific polypeptides. In a specific example, the antibodies interact with the breast, for example, breast cancer cells, and fluoresce upon contact such that imaging and visibility of the breast are enhanced to allow a determination of the diseased or non-diseased state of the breast.

The present invention will be further described with reference to the following examples; however, it is to be understood that the present invention is not limited to such examples. All parts or amounts, unless otherwise specified, are by weight.

In order to facilitate understanding of the following examples certain frequently occurring methods and/or terms will be described.

"Plasmids" are designated by a lower case p preceded and/or followed by capital letters and/or numbers. The starting plasmids herein are either commercially available, publicly available on an unrestricted basis, or can be constructed from available plasmids in accord with published procedures. In addition, equivalent plasmids to those

described are known in the art and will be apparent to the ordinarily skilled artisan.

"Digestion" of DNA refers to catalytic cleavage of the DNA with a restriction enzyme that acts only at certain sequences in the DNA. The various restriction enzymes used herein are commercially available and their reaction conditions, cofactors and other requirements were used as would be known to the ordinarily skilled artisan. For analytical purposes, typically 1  $\mu\text{g}$  of plasmid or DNA fragment is used with about 2 units of enzyme in about 20  $\mu\text{l}$  of buffer solution. For the purpose of isolating DNA fragments for plasmid construction, typically 5 to 50  $\mu\text{g}$  of DNA are digested with 20 to 250 units of enzyme in a larger volume. Appropriate buffers and substrate amounts for particular restriction enzymes are specified by the manufacturer. Incubation times of about 1 hour at 37°C are ordinarily used, but may vary in accordance with the supplier's instructions. After digestion the reaction is electrophoresed directly on a polyacrylamide gel to isolate the desired fragment.

Size separation of the cleaved fragments is performed using 1 percent TAE agarose gel described by Sambrook, et al., "Molecular Cloning: A Laboratory Manual" Cold Spring Laboratory Press, (1989).

"Oligonucleotides" refers to either a single stranded polydeoxynucleotide or two complementary polydeoxynucleotide strands which may be chemically synthesized. Such synthetic oligonucleotides have no 5' phosphate and thus will not ligate to another oligonucleotide without adding a phosphate with an ATP in the presence of a kinase. A synthetic oligonucleotide will ligate to a fragment that has not been dephosphorylated.

"Ligation" refers to the process of forming phosphodiester bonds between two double stranded nucleic acid fragments (Maniatis, T., et al., Id., p. 146). Unless

otherwise provided, ligation may be accomplished using known buffers and conditions with 10 units of T4 DNA ligase ("ligase") per 0.5  $\mu$ g of approximately equimolar amounts of the DNA fragments to be ligated.

Unless otherwise stated, transformation was performed as described in the method of Graham, F. and Van der Eb, A., Virology, 52:456-457 (1973).

#### Example 1

##### Determination of Transcription of a breast specific gene

To assess the presence or absence of active transcription of a breast specific gene RNA, approximately 6 ml of venous blood is obtained with a standard venipuncture technique using heparinized tubes. Whole blood is mixed with an equal volume of phosphate buffered saline, which is then layered over 8 ml of Ficoll (Pharmacia, Uppsala, Sweden) in a 15-ml polystyrene tube. The gradient is centrifuged at 1800 X g for 20 min at 5°C. The lymphocyte and granulocyte layer (approximately 5 ml) is carefully aspirated and rediluted up to 50 ml with phosphate-buffered saline in a 50-ml tube, which is centrifuged again at 1800 X g for 20 min. at 5°C. The supernatant is discarded and the pellet containing nucleated cells is used for RNA extraction using the RNazole B method as described by the manufacturer (Tel-Test Inc., Friendswood, TX).

To determine the quantity of mRNA from the gene of interest, a probe is designed with an identity to at least a portion of the mRNA sequence transcribed from a human gene whose coding portion includes a DNA sequence of Figures 1-20 (SEQ ID NO:1-20). This probe is mixed with the extracted RNA and the mixed DNA and RNA are precipitated with ethanol -70°C for 15 minutes). The pellet is resuspended in hybridization buffer and dissolved. The tubes containing the mixture are incubated in a 72°C water bath for 10-15 mins. to denature the DNA. The tubes are rapidly transferred to a water bath

at the desired hybridization temperature. Hybridization temperature depends on the G + C content of the DNA. Hybridization is done for 3 hrs. 0.3 ml of nuclease-S1 buffer is added and mixed well. 50  $\mu$ l of 4.0 M ammonium acetate and 0.1 M EDTA is added to stop the reaction. The mixture is extracted with phenol/chloroform and 20  $\mu$ g of carrier tRNA is added and precipitation is done with an equal volume of isopropanol. The precipitate is dissolved in 40  $\mu$ l of TE (pH 7.4) and run on an alkaline agarose gel. Following electrophoresis, the RNA is microsequenced to confirm the nucleotide sequence. (See Favalaro, J. et al., Methods Enzymol., 65:718 (1980) for a more detailed review).

Two oligonucleotide primers are employed to amplify the sequence isolated by the above methods. The 5' primer is 20 nucleotides long and the 3' primer is a complimentary sequence for the 3' end of the isolated mRNA. The primers are custom designed according to the isolated mRNA. The reverse transcriptase reaction and PCR amplification are performed sequentially without interruption in a Perkin Elmer 9600 PCR machine (Emeryville, CA). Four hundred ng total RNA in 20  $\mu$ l diethylpyrocarbonate-treated water are placed in a 65°C water bath for 5 min. and then quickly chilled on ice immediately prior to the addition of PCR reagents. The 50- $\mu$ l total PCR volume consisted of 2.5 units Taq polymerase (Perkin-Elmer). 2 units avian myeloblastosis virus reverse transcriptase (Boehringer Mannheim, Indianapolis, IN); 200  $\mu$ M each of dCTP, dATP, dGTP and dTTP (Perkin Elmer); 18 pM each primer, 10 mM Tris-HCl; 50 mM KCl; and 2 mM MgCl<sub>2</sub> (Perkin Elmer). PCR conditions are as follows: cycle 1 is 42°C for 15 min then 97°C for 15 s (1 cycle); cycle 2 is 95°C for 1 min. 60°C for 1 min, and 72°C for 30 s (15 cycles); cycle 3 is 95°C for 1 min. 60°C for 1 min., and 72°C for 1 min. (10 cycles); cycle 4 is 95°C for 1 min., 60°C for 1 min., and 72°C for 2 min. (8 cycles); cycle 5 is 72°C for 15 min. (1 cycle); and the final cycle is a 4°C hold until sample is

taken out of the machine. The 50- $\mu$ l PCR products are concentrated down to 10  $\mu$ l with vacuum centrifugation, and a sample is then run on a thin 1.2 % Tris-borate-EDTA agarose gel containing ethidium bromide. A band of expected size would indicate that this gene is present in the tissue assayed. The amount of RNA in the pellet may be quantified in numerous ways, for example, it may be weighed.

Verification of the nucleotide sequence of the PCR products is done by microsequencing. The PCR product is purified with a Qiagen PCR Product Purification Kit (Qiagen, Chatsworth, CA) as described by the manufacturer. One  $\mu$ g of the PCR product undergoes PCR sequencing by using the Taq DyeDeoxy Terminator Cycle sequencing kit in a Perkin-Elmer 9600 PCR machine as described by Applied Biosystems (Foster, CA). The sequenced product is purified using Centri-Sep columns (Princeton Separations, Adelphia, NJ) as described by the company. This product is then analyzed with an ABI model 373A DNA sequencing system (Applied Biosystems) integrated with a Macintosh IIci computer.

#### Example 2

##### Bacterial Expression and Purification of the BSG Proteins and Use For Preparing a Monoclonal Antibody

The DNA sequence encoding a polypeptide of the present invention, for this example BSG1, ATCC # 97175, is initially amplified using PCR oligonucleotide primers corresponding to the 5' sequences of the protein and the vector sequences 3' to the protein. Additional nucleotides corresponding to the DNA sequence are added to the 5' and 3' sequences respectively. The 5' oligonucleotide primer has the sequence 5' GCCACCATGGATGTTTTCAAG 3' (SEQ ID NO:21) and contains an NcoI restriction enzyme site followed by 15 nucleotides of coding sequence starting from the initial amino acid of the processed protein. The 3' sequence 5' GCGCAGATCTGTCT

CCCCCACTCTGGGC 3' (SEQ ID NO:22) and contains a complementary sequence to a BglIII restriction enzyme site and is followed by 18 nucleotides of the nucleic acid sequence encoding the protein. The restriction enzyme sites correspond to the restriction enzyme sites on a bacterial expression vector, pQE-60 (Qiagen, Inc. Chatsworth, CA). pQE-60 encodes antibiotic resistance (Amp<sup>r</sup>), a bacterial origin of replication (ori), an IPTG-regulatable promoter operator (P/O), a ribosome binding site (RBS), a 6-His tag and restriction enzyme sites. pQE-60 is then digested with NcoI and BglIII. The amplified sequences are ligated into pQE-60 and inserted in frame with the sequence encoding for the histidine tag and the RBS. The ligation mixture is then used to transform an *E. coli* strain M15/rep 4 (Qiagen) by the procedure described in Sambrook, J. et al., Molecular Cloning: A Laboratory Manual, Cold Spring Laboratory Press, (1989). M15/rep4 contains multiple copies of the plasmid pREP4, which expresses the lacI repressor and also confers kanamycin resistance (Kan<sup>r</sup>). Transformants are identified by their ability to grow on LB plates and ampicillin/kanamycin resistant colonies are selected. Plasmid DNA is isolated and confirmed by restriction analysis.

Clones containing the desired constructs are grown overnight (O/N) in liquid culture in LB media supplemented with both Amp (100 ug/ml) and Kan (25 ug/ml). The O/N culture is used to inoculate a large culture at a ratio of 1:100 to 1:250. The cells are grown to an optical density 600 (O.D.<sup>600</sup>) of between 0.4 and 0.6. IPTG ("Isopropyl-B-D-thiogalacto pyranoside") is then added to a final concentration of 1 mM. IPTG induces by inactivating the lacI repressor, clearing the P/O leading to increased gene expression. Cells are grown an extra 3 to 4 hours. Cells are then harvested by centrifugation. The cell pellet is solubilized in the chaotropic agent 6 Molar Guanidine HCl. After clarification, solubilized protein is purified from

this solution by chromatography on a Nickel-Chelate column under conditions that allow for tight binding by proteins containing the 6-His tag (Hochuli, E. et al., J. Chromatography 411:177-184 (1984)). BSG1 protein (>90% pure) is eluted from the column in 6 molar guanidine HCl pH 5.0 and for the purpose of renaturation adjusted to 3 molar guanidine HCl, 100mM sodium phosphate, 10 mmolar glutathione (reduced) and 2 mmolar glutathione (oxidized). After incubation in this solution for 12 hours the protein is dialyzed to 10 mmolar sodium phosphate.

The protein purified in this manner may be used as an epitope to raise monoclonal antibodies specific to such protein. The monoclonal antibodies generated against the polypeptide the isolated protein can be obtained by direct injection of the polypeptides into an animal or by administering the polypeptides to an animal. The antibodies so obtained will then bind to the protein itself. Such antibodies can then be used to isolate the protein from tissue expressing that polypeptide by the use of an, for example, ELISA assay.

### Example 3

#### Preparation of cDNA Libraries from Breast Tissue

Total cellular RNA is prepared from tissues by the guanidinium-phenol method as previously described (P. Chomczynski and N. Sacchi, Anal. Biochem., 162: 156-159 (1987)) using RNazol (Cinna-Biotech). An additional ethanol precipitation of the RNA is included. Poly A mRNA is isolated from the total RNA using oligo dT-coated latex beads (Qiagen). Two rounds of poly A selection are performed to ensure better separation from non-polyadenylated material when sufficient quantities of total RNA are available.

The mRNA selected on the oligo dT is used for the synthesis of cDNA by a modification of the method of Gobbler and Hoffman (Gobbler, U. and B.J. Hoffman, 1983, Gene,



25:263). The first strand synthesis is performed using either Moloney murine sarcoma virus reverse transcriptase (Stratagene) or Superscript II (RNase H minus Moloney murine reverse transcriptase, Gibco-BRL). First strand synthesis is primed using a primer/linker containing an Xho I restriction site. The nucleotide mix used in the synthesis contains methylated dCTP to prevent restriction within the cDNA sequence. For second-strand synthesis *E. coli* polymerase Klenow fragment is used and [<sup>32</sup>P]-dATP is incorporated as a tracer of nucleotide incorporation.

Following 2nd strand synthesis, the cDNA is made blunt ended using either T4 DNA polymerase or Klenow fragment. Eco RI adapters are added to the cDNA and the cDNA is restricted with Xho I. The cDNA is size fractionated over a Sephacryl S-500 column (Pharmacia) to remove excess linkers and cDNAs under approximately 500 base pairs.

The cDNA is cloned unidirectionally into the Eco RI-Xho I sites of either pBluescript II phagemid or lambda Uni-zap XR (Stratagene). In the case of cloning into pBluescript II, the plasmids are electroporated into E.coli SURE competent cells (Stratagene). When the cDNA is cloned into Uni-Zap XR it is packaged using the Gigipack II packaging extract (Stratagene). The packaged phage is used to infect SURE cells and amplified. The pBluescript phagemid containing the cDNA inserts are excised from the lambda Zap phage using the helper phage ExAssist (Stratagene). The rescued phagemid is plated on SOLR E.coli cells (Stratagene).

#### Preparation of Sequencing Templates

Template DNA for sequencing is prepared by 1) a boiling method or 2) PCR amplification.

The boiling method is a modification of the method of Holmes and Quigley (Holmes, D.S. and M. Quigley, 1981, Anal. Biochem., 114:193). Colonies from either cDNA cloned into Bluescript II or rescued Bluescript phagemid are grown in an enriched bacterial media overnight. 400 µl of cells are

centrifuged and resuspended in STET (0.1M NaCl, 10mM TRIS Ph 8.0, 1.0 mM EDTA and 5% Triton X-100) including lysozyme (80 µg/ml) and RNase A (4 µg/ml). Cells are boiled for 40 seconds and centrifuged for 10 minutes. The supernatant is removed and the DNA is precipitated with PEG/NaCl and washed with 70% ethanol (2x). Templates are resuspended in water at approximately 250 ng/µl.

Preparation of templates by PCR is a modification of the method of Rosenthal et al. (Rosenthal, et al., Nucleic Acids Res., 1993, 21:173-174). Colonies containing cDNA cloned into pBluescript II or rescued pBluescript phagemid are grown overnight in LB containing ampicillin in a 96 well tissue culture plate. Two µl of the cultures are used as template in a PCR reaction (Saiki, RK, et al., Science, 239:487-493, 1988; and Saiki, RK, et al., Science, 230:1350-1354, 1985) using a tricine buffer system (Ponce and Micol., Nucleic Acids Res., 1992, 20:1992.) and 200 µM dNTPs. The primer set chosen for amplification of the templates is outside of primer sites chosen for sequencing of the templates. The primers used are 5'-ATGCTTCCGGCTCGTATG-3' (SEQ ID NO:23) which is 5' of the M13 reverse sequence in pBluescript and 5'-GGGTTTTCACGTCACGAC-3' (SEQ ID NO:24) which is 3' of the M13 forward primer in pBluescript. Any primers which correspond to the sequence flanking the M13 forward and reverse sequences can be used. Perkin-Elmer 9600 thermocyclers are used for amplification of the templates with the following cycler conditions: 5 min at 94°C (1 cycle); (20 sec at 94°C); 20 sec at 55°C (1 min at 72°C) (30 cycles); 7 min at 72°C (1 cycle). Following amplification the PCR templates are precipitated using PEG/NaCl and washed three times with 70% ethanol. The templates are resuspended in water.

#### Example 4

#### Isolation of a Selected Clone From Breast Tissue

Two approaches are used to isolate a particular clone from a cDNA library prepared from human breast tissue.

In the first, a clone is isolated directly by screening the library using an oligonucleotide probe. To isolate a particular clone, a specific oligonucleotide with 30-40 nucleotides is synthesized using an Applied Biosystems DNA synthesizer according to one of the partial sequences described in this application. The oligonucleotide is labeled with  $^{32}\text{P}$ -ATP using T4 polynucleotide kinase and purified according to the standard protocol (Maniatis et al., Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Press, Cold Spring, NY, 1982). The Lambda cDNA library is plated on 1.5% agar plate to a density of 20,000-50,000 pfu/150 mm plate. These plates are screened using Nylon membranes according to the standard phage screening protocol (Stratagene, 1993). Specifically, the Nylon membrane with denatured and fixed phage DNA is prehybridized in 6 x SSC, 20 mM  $\text{NaH}_2\text{PO}_4$ , 0.4% SDS, 5 x Denhardt's 500  $\mu\text{g/ml}$  denatured, sonicated salmon sperm DNA; and 6 x SSC, 0.1% SDS. After one hour of prehybridization, the membrane is hybridized with hybridization buffer 6 x SSC, 20 mM  $\text{NaH}_2\text{PO}_4$ , 0.4% SDS, 500  $\mu\text{g/ml}$  denatured, sonicated salmon sperm DNA with  $1 \times 10^6$  cpm/ml  $^{32}\text{P}$ -probe overnight at 42°C. The membrane is washed at 45-50°C with washing buffer 6 x SSC, 0.1% SDS for 20-30 minutes dried and exposed to Kodak X-ray film overnight. Positive clones are isolated and purified by secondary and tertiary screening. The purified clone sequenced to verify its identity to the partial sequence described in this application.

An alternative approach to screen the cDNA library prepared from human breast tissue is to prepare a DNA probe corresponding to the entire partial sequence. To prepare a probe, two oligonucleotide primers of 17-20 nucleotides derived from both ends of the partial sequence reported are synthesized and purified. These two oligonucleotides are

used to amplify the probe using the cDNA library template. The DNA template is prepared from the phage lysate of the cDNA library according to the standard phage DNA preparation protocol (Maniatis et al.). The polymerase chain reaction is carried out in 25  $\mu$ l reaction mixture with 0.5  $\mu$ g of the above cDNA template. The reaction mixture is 1.5-5 mM  $MgCl_2$ , 0.01% (w/v) gelatin, 20  $\mu$ M each of dATP, dCTP, dGTP, dTTP, 25 pmol of each primer and 0.25 Unit of Taq polymerase. Thirty five cycles of PCR (denaturation at 94°C for 1 min; annealing at 55°C for 1 min; elongation at 72°C for 1 min) are performed with the Perkin-Elmer Cetus automated thermal cycler. The amplified product is analyzed by agarose gel electrophoresis and the DNA band with expected molecular weight is excised and purified. The PCR product is verified to be the probe by subcloning and sequencing the DNA product. The probe is labeled with the Multiprime DNA Labelling System (Amersham) at a specific activity  $< 1 \times 10^9$  dmp/ $\mu$ g. This probe is used to screen the lambda cDNA library according to Stratagene's protocol. Hybridization is carried out with 5X TEN 920XTEN:0.3M Tris-HCl pH 8.0, 0.02M EDTA and 3MNaCl), 5X Denhardt's, 0.5% sodium pyrophosphate, 0.1% SDS, 0.2 mg/ml heat denatured salmon sperm DNA and  $1 \times 10^6$  cpm/ml of [ $^{32}P$ ]-labeled probe at 55°C for 12 hours. The filters are washed in 0.5X TEN at room temperature for 20-30 min., then at 55°C for 15 min. The filters are dried and autoradiographed at -70°C using Kodak XAR-5 film. The positive clones are purified by secondary and tertiary screening. The sequence of the isolated clone are verified by DNA sequencing.

General procedures for obtaining complete sequences from partial sequences described herein are summarized as follows;  
**Procedure 1**

Selected human DNA from the partial sequence clone (the cDNA clone that was sequenced to give the partial sequence) is purified e.g., by endonuclease digestion using Eco-R1, gel electrophoresis, and isolation of the clone by removal from

low melting agarose gel. The isolated insert DNA, is radiolabeled e.g., with  $^{32}\text{P}$  labels, preferably by nick translation or random primer labeling. The labeled insert is used as a probe to screen a lambda phage cDNA library or a plasmid cDNA library. Colonies containing clones related to the probe cDNA are identified and purified by known purification methods. The ends of the newly purified clones are nucleotide sequenced to identify full length sequences. Complete sequencing of full length clones is then performed by Exonuclease III digestion or primer walking. Northern blots of the mRNA from various tissues using at least part of the deposited clone from which the partial sequence is obtained as a probe can optionally be performed to check the size of the mRNA against that of the purported full length cDNA.

The following procedures 2 and 3 can be used to obtain full length genes or full length coding portions of genes where a clone isolated from the deposited clone mixture does not contain a full length sequence. A library derived from human breast tissue or from the deposited clone mixture is also applicable to obtaining full length sequences from clones obtained from sources other than the deposited mixture by use of the partial sequences of the present invention.

#### **Procedure 2**

##### **RACE Protocol For Recovery of Full-Length Genes**

Partial cDNA clones can be made full-length by utilizing the rapid amplification of cDNA ends (RACE) procedure described in Frohman, M.A., Dush, M.K. and Martin, G.R. (1988) Proc. Nat'l. Acad. Sci. USA, 85:8998-9002. A cDNA clone missing either the 5' or 3' end can be reconstructed to include the absent base pairs extending to the translational start or stop codon, respectively. In most cases, cDNAs are missing the start of translation therefor. The following briefly describes a modification of this original 5' RACE

procedure. Poly A+ or total RNA is reverse transcribed with Superscript II (Gibco/BRL) and an antisense or complementary primer specific to the cDNA sequence. The primer is removed from the reaction with a Microcon Concentrator (Amicon). The first-strand cDNA is then tailed with dATP and terminal deoxynucleotide transferase (Gibco/BRL). Thus, an anchor sequence is produced which is needed for PCR amplification. The second strand is synthesized from the dA-tail in PCR buffer, Taq DNA polymerase (Perkin-Elmer Cetus), an oligo-dT primer containing three adjacent restriction sites (XhoI, SalI and ClaI) at the 5' end and a primer containing just these restriction sites. This double-stranded cDNA is PCR amplified for 40 cycles with the same primers as well as a nested cDNA-specific antisense primer. The PCR products are size-separated on an ethidium bromide-agarose gel and the region of gel containing cDNA products the predicted size of missing protein-coding DNA is removed. cDNA is purified from the agarose with the Magic PCR Prep kit (Promega), restriction digested with XhoI or SalI, and ligated to a plasmid such as pBluescript SKII (Stratagene) at ShoI and EcoRV sites. This DNA is transformed into bacteria and the plasmid clones sequenced to identify the correct protein-coding inserts. Correct 5' ends are confirmed by comparing this sequence with the putatively identified homologue and overlap with the partial cDNA clone.

Several quality-controlled kits are available for purchase. Similar reagents and methods to those above are supplied in kit form from Gibco/BRL. A second kit is available from Clontech which is a modification of a related technique, SLIC (single-stranded ligation to single-stranded cDNA) developed by Dumas et al. (Dumas, J.B., Edwards, M., Delort, J. and Mallet, Jr., 1991, Nucleic Acids Res., 19:5227-5232). The major differences in procedure are that the RNA is alkaline hydrolyzed after reverse transcription and RNA ligase is used to join a restriction site-containing

anchor primer to the first-strand cDNA. This obviates the necessity for the dA-tailing reaction which results in a polyT stretch that is difficult to sequence past.

An alternative to generating 5' cDNA from RNA is to use cDNA library double-stranded DNA. An asymmetric PCR-amplified antisense cDNA strand is synthesized with an antisense cDNA-specific primer and a plasmid-anchored primer. These primers are removed and a symmetric PCR reaction is performed with a nested cDNA-specific antisense primer and the plasmid-anchored primer.

### Procedure 3

#### RNA Ligase Protocol For Generating The 5' End Sequences To Obtain Full Length Genes

Once a gene of interest is identified, several methods are available for the identification of the 5' or 3' portions of the gene which may not be present in the original deposited clone. These methods include but are not limited to filter probing, clone enrichment using specific probes and protocols similar and identical to 5' and 3' RACE. While the full length gene may be present in a library and can be identified by probing, a useful method for generating the 5' end is to use the existing sequence information from the original partial sequence to generate the missing information. A method similar to 5' RACE is available for generating the missing 5' end of a desired full-length gene. (This method was published by Fromont-Racine et al, Nucleic Acids Res., 21(7):1683-1684 (1993). Briefly, a specific RNA oligonucleotide is ligated to the 5' ends of a population of RNA presumably containing full-length gene RNA transcript and a primer set containing a primer specific to the ligated RNA oligonucleotide. A primer specific to a known sequence (EST) of the gene of interest is used to PCR amplify the 5' portion of the desired full length gene which may then be sequenced and used to generate the full length gene. This method

starts with total RNA isolated from the desired source, poly A RNA may be used but is not a prerequisite for this procedure. The RNA preparation may then be treated with phosphatase if necessary to eliminate 5' phosphate groups on degraded or damaged RNA which may interfere with the later RNA ligase step. The phosphatase if used is then inactivated and the RNA is treated with tobacco acid pyrophosphatase in order to remove the cap structure present at the 5' ends of messenger RNAs. This reaction leaves a 5' phosphate group at the 5' end of the cap-cleaved RNA which can then be ligated to an RNA oligonucleotide using T4 RNA ligase. This modified RNA preparation can then be used as a template for first strand cDNA synthesis using a gene-specific oligonucleotide. The first strand synthesis reaction can then be used as a template for PCR amplification of the desired 5' end using a primer specific to the ligated RNA oligonucleotide and a primer specific to the known sequence (EST) of the gene of interest. The resultant product is then sequenced and analyzed to confirm that the 5' end sequence belongs to the partial sequence.

#### Example 5

##### Cloning and expression of BSG1 using the baculovirus expression system

The DNA sequence encoding the full length BSG1 protein, ATCC # 97175, was amplified using PCR oligonucleotide primers corresponding to the 5' and 3' sequences of the gene:

The 5' primer has the sequence 5' AAAGGATCCCCCGCCATCATGG ATGTTTTCAAGAAG 3' (SEQ ID NO:25) and contains a BamHI restriction enzyme site (in bold) followed by 8 nucleotides resembling an efficient signal for the initiation of translation in eukaryotic cells (Kozak, M., J. Mol. Biol., 196:947-950 (1987) of the BSG1 gene (the initiation codon for translation "ATG" is underlined).

The 3' primer has the sequence 5' AAATCTAGACTAGTCTCCCCC



ACTCTG 3' (SEQ ID NO:26) and contains the cleavage site for the restriction endonuclease XbaI and 21 nucleotides complementary to the 3' sequence of the BSG1 gene. The amplified sequences were isolated from a 1% agarose gel using a commercially available kit ("Geneclean," BIO 101 Inc., La Jolla, Ca.). The fragment was then digested with the endonucleases BamHI and XbaI and then purified again on a 1% agarose gel. This fragment is designated F2.

The vector pA2 (modification of pVL941 vector, discussed below) is used for the expression of the BSG1 protein using the baculovirus expression system (for review see: Summers, M.D. and Smith, G.E. 1987, A manual of methods for baculovirus vectors and insect cell culture procedures, Texas Agricultural Experimental Station Bulletin No. 1555). This expression vector contains the strong polyhedrin promoter of the Autographa californica nuclear polyhedrosis virus (AcMNPV) followed by the recognition sites for the restriction endonucleases BamHI and XbaI. The polyadenylation site of the simian virus (SV)40 is used for efficient polyadenylation. For an easy selection of recombinant virus the beta-galactosidase gene from E.coli is inserted in the same orientation as the polyhedrin promoter followed by the polyadenylation signal of the polyhedrin gene. The polyhedrin sequences are flanked at both sides by viral sequences for the cell-mediated homologous recombination of co-transfected wild-type viral DNA. Many other baculovirus vectors could be used in place of pA2 such as pRG1, pAc373, pVL941 and pAcIM1 (Luckow, V.A. and Summers, M.D., Virology, 170:31-39).

The plasmid was digested with the restriction enzymes BamHI and XbaI and dephosphorylated using calf intestinal phosphatase by procedures known in the art. The DNA was then isolated from a 1% agarose gel using the commercially available kit ("Geneclean" BIO 101 Inc., La Jolla, Ca.). This vector DNA is designated V2.

Fragment F2 and the dephosphorylated plasmid pA2 were ligated with T4 DNA ligase. *E.coli* HB101 cells were then transformed and bacteria identified that contained the plasmid (pBacBSG1) with the BSG1 gene using the enzymes BamHI and XbaI. The sequence of the cloned fragment was confirmed by DNA sequencing.

5  $\mu$ g of the plasmid pBacBSG1 was co-transfected with 1.0  $\mu$ g of a commercially available linearized baculovirus ("BaculoGold™ baculovirus DNA", Pharmingen, San Diego, CA.) using the lipofection method (Felgner et al. Proc. Natl. Acad. Sci. USA, 84:7413-7417 (1987)).

1 $\mu$ g of BaculoGold™ virus DNA and 5  $\mu$ g of the plasmid pBacBSG1 were mixed in a sterile well of a microtiter plate containing 50  $\mu$ l of serum free Grace's medium (Life Technologies Inc., Gaithersburg, MD). Afterwards 10  $\mu$ l Lipofectin plus 90  $\mu$ l Grace's medium were added, mixed and incubated for 15 minutes at room temperature. Then the transfection mixture was added drop-wise to the Sf9 insect cells (ATCC CRL 1711) seeded in a 35 mm tissue culture plate with 1 ml Grace's medium without serum. The plate was rocked back and forth to mix the newly added solution. The plate was then incubated for 5 hours at 27°C. After 5 hours the transfection solution was removed from the plate and 1 ml of Grace's insect medium supplemented with 10% fetal calf serum was added. The plate was put back into an incubator and cultivation continued at 27°C for four days.

After four days the supernatant was collected and a plaque assay performed similar as described by Summers and Smith (supra). As a modification an agarose gel with "Blue Gal" (Life Technologies Inc., Gaithersburg) was used which allows an easy isolation of blue stained plaques. (A detailed description of a "plaque assay" can also be found in the user's guide for insect cell culture and baculovirology distributed by Life Technologies Inc., Gaithersburg, page 9-10).

Four days after the serial dilution, the virus was added to the cells and blue stained plaques were picked with the tip of an Eppendorf pipette. The agar containing the recombinant viruses was then resuspended in an Eppendorf tube containing 200  $\mu$ l of Grace's medium. The agar was removed by a brief centrifugation and the supernatant containing the recombinant baculovirus was used to infect Sf9 cells seeded in 35 mm dishes. Four days later the supernatants of these culture dishes were harvested and then stored at 4°C.

Sf9 cells were grown in Grace's medium supplemented with 10% heat-inactivated FBS. The cells were infected with the recombinant baculovirus V-BSG1 at a multiplicity of infection (MOI) of 2. Six hours later the medium was removed and replaced with SF900 II medium minus methionine and cysteine (Life Technologies Inc., Gaithersburg). 42 hours later 5  $\mu$ Ci of  $^{35}$ S-methionine and 5  $\mu$ Ci  $^{35}$ S cysteine (Amersham) were added. The cells were further incubated for 16 hours before they were harvested by centrifugation and the labelled proteins visualized by SDS-PAGE and autoradiography.

#### Example 6

##### Expression of Recombinant BSG1 in COS cells

The expression of plasmid, BSG1 HA is derived from a vector pcDNA1/Amp (Invitrogen) containing: 1) SV40 origin of replication, 2) ampicillin resistance gene, 3) E.coli replication origin, 4) CMV promoter followed by a polylinker region, an SV40 intron and polyadenylation site. A DNA fragment encoding the entire precursor and a HA tag fused in frame to its 3' end was cloned into the polylinker region of the vector, therefore, the recombinant protein expression is directed under the CMV promoter. The HA tag corresponds to an epitope derived from the influenza hemagglutinin protein as previously described (I. Wilson, H. Niman, R. Heighten, A. Cherenson, M. Connolly, and R. Lerner, 1984, Cell 37:767, (1984)). The infusion of HA tag to the target protein allows

easy detection of the recombinant protein with an antibody that recognizes the HA epitope.

The plasmid construction strategy is described as follows:

The DNA sequence encoding BSG1, ATCC # 97175, was constructed by PCR using two primers: the 5' primer AAAGGA TCCCCCGCCATCATGGATGTTTTCAAGAAG 3' (SEQ ID NO:27) contains a BamHI site followed by 18 nucleotides of BSG1 coding sequence starting from the initiation codon; the 3' sequence AAATC TAGACTAAAGCGTAGTCTGGGACGTCGTATGGGTACTCCTGGGGTCTCCCCCACTCTGGGC 3' (SEQ ID NO:28) contains complementary sequences to an XbaI site, translation stop codon, HA tag and the last 18 nucleotides of the BamHI coding sequence (not including the stop codon). Therefore, the PCR product contains an BamHI site, BSG1 coding sequence followed by HA tag fused in frame, a translation termination stop codon next to the HA tag, and an XbaI site. The PCR amplified DNA fragment and the vector, pcDNA1/Amp, were digested with BamHI and XbaI restriction enzyme and ligated. The ligation mixture was transformed into E. coli strain SURE (available from Stratagene Cloning Systems, 11099 North Torrey Pines Road, La Jolla, CA 92037) the transformed culture was plated on ampicillin media plates and resistant colonies were selected. Plasmid DNA was isolated from transformants and examined by restriction analysis for the presence of the correct fragment. For expression of the recombinant BSG protein, COS cells were transfected with the expression vector by DEAE-DEXTRAN method (J. Sambrook, E. Fritsch, T. Maniatis, Molecular Cloning: A Laboratory Manual, Cold Spring Laboratory Press, (1989)). The expression of the BSG HA protein was detected by radiolabelling and immunoprecipitation method (E. Harlow, D. Lane, Antibodies: A Laboratory Manual, Cold Spring Harbor Laboratory Press, (1988)). Cells were labelled for 8 hours with <sup>35</sup>S-cysteine two days post transfection. Culture media was then collected and cells were lysed with detergent (RIPA

buffer (150 mM NaCl, 1% NP-40, 0.1% SDS, 1% NP-40, 0.5% DOC, 50mM Tris, pH 7.5) (Wilson, I. et al., Id. 37:767 (1984)). Both cell lysate and culture media were precipitated with an HA specific monoclonal antibody. Proteins precipitated were analyzed on 15% SDS-PAGE gels.

Numerous modifications and variations of the present invention are possible in light of the above teachings and, therefore, within the scope of the appended claims, the invention may be practiced otherwise than as particularly described.

WHAT IS CLAIMED IS:

1. An isolated polynucleotide comprising a member selected from the group consisting of
  - (a) a polynucleotide encoding the same polypeptide as the polynucleotide of Figure 1 (SEQ ID NO:1);
  - (b) a polynucleotide encoding the same mature polypeptide as a human gene having a coding portion which includes DNA having at least a 90% identity to the DNA of one of Figures 2-20 (SEQ ID NO:2-20);
  - (c) a polynucleotide which hybridizes to the polynucleotide of (a) and which has at least a 70% identity thereto; and
  - (d) a polynucleotide encoding the same mature polypeptide as a human gene having a coding portion which includes DNA having at least a 90% identity to a DNA included in the deposited clone.
2. The polynucleotide of Claim 1 wherein the human gene includes DNA contained in the deposited clone.
3. The polynucleotide of Claim 1 wherein the member is a polynucleotide encoding the same polypeptide as the polynucleotide of Figure 1 (SEQ ID NO:1).
4. A vector containing the polynucleotide of claim 1.
5. A host cell transformed or transfected with the vector of Claim 4.
6. A process for producing cells capable of expressing a polypeptide comprising genetically engineering cells with the vector of Claim 4.

7. A process for producing a polypeptide comprising: expressing from the host cell of Claim 5 the polypeptide encoded by said polynucleotide.

8. A polypeptide comprising a member selected from the group consisting of: (i) a polypeptide encoded by a human gene, said human gene having a coding portion whose DNA has at least a 90% identity to the DNA of one of Figures 2-20 (SEQ ID NO:2-20); (ii) a polypeptide having the deduced amino acid sequence as set forth in Figure 1 (SEQ ID NO:1) and fragments, analogs and derivatives thereof; and (iii) a polypeptide encoded by the human gene whose coding region includes a DNA having at least a 90% identity to the DNA contained in the deposited clone and fragments, analogs and derivatives of said polypeptide.

9. The polypeptide of Claim 8 wherein the polypeptide has the deduced amino acid sequence as set forth in Figure 1 (SEQ ID NO:1).

10. An antibody against the polypeptide of claim 8.

11. A compound which inhibits activation of the polypeptide of claim 8.

12. A method for the treatment of a patient having need to inhibit a breast specific gene protein comprising: administering to the patient a therapeutically effective amount of the compound of Claim 11.

13. The method of claim 12 wherein the compound is a polypeptide and the therapeutically effective amount of the compound is administered by providing to the patient DNA encoding said polypeptide and expressing said polypeptide in vivo.

14. A method for the treatment of a patient having need of a breast specific gene protein comprising: administering to the patient a therapeutically effective amount of the polypeptide of claim 8.

15. A process for diagnosing a disorder of the breast in a host comprising:

determining transcription of a human gene in a sample derived from non-breast tissue of a host, said gene having a coding portion which includes DNA having at least 90% identity to DNA selected from the group consisting of the DNA of Figures 1-20 (SEQ ID NO:1-20), whereby said transcription indicates a disorder of the breast in the host.

16. The process of claim 15 wherein transcription is determined by detecting the presence of an altered level of RNA transcribed from said human gene.

17. The process of claim 15 wherein transcription is determined by detecting the presence of an altered level of DNA complementary to the RNA transcribed from said human gene.

18. The process of claim 15 wherein transcription is determined by detecting the presence of an altered level of an expression product of said human gene.

19. A process for determining a disorder of a breast in a host comprising:

contacting an antibody specific to a BSG antigen or an epitopic portion thereof, to a fluid sample derived from a host;

determining the presence of an altered level of a BSG gene product in said sample.



20. A process for identifying antagonists to the polypeptide of claim 8 comprising:

contacting said polypeptide with a natural substrate and a labeled compound to be screened either simultaneously or in either consecutive order; and

determining whether the therapeutic effectively competes with the natural substrate in a manner sufficient to prevent binding of the protein to its substrate.

1/14

## FIG. 1

## BSG1 (assembly sequence)

GCCGCTGCGCAGACTCGAGCCAGCTCAAGCCCGCAGCTCGCAGGGAGAT  
CCAGCTCCGTCTGCTGCAGCAGCMCAACCTGTCACACCCACCATGGAT  
GTYTTCAAGAAGGGCTTCTCCATCGCCAAGGAGGGCGTGGTGCGGTG  
GAAAAGACCAAGCAGGGGTGACGGAAGCAGCTGAGAAGACCAAGGAGGG  
GGTCATGTATGTGGAGCCCAAGACCAAGGAGATGTTGTATGTACAGAGCGT  
GACCTCAGTGGCCGAGAACCAAGGAGCAGGCCAACGCCGTGAGCAAGG  
CTGTGGTGAGCAGCGTCAACACTKTGGCCACCAAGACCGTKGAGGAGGCGG  
AGAACATCGCGGTCACTCCGGGTGKTGCGCAAGGAGGAYTTKAGGCCAT  
YTKCCCCCAACAGGAGGTGAGGCATCMARAGARARGAKWGSAAAGWGGC  
MRAKRRGGMSCAGAGTGGGGAGACTAGAGGGCTACAGGCCAGCTTGGAT  
GACCTGAAGAGCGCTCCTCTGCCTTGGGACACCATCCCTCCTAGCACAAAG  
GAGTGCCCGCTTGTAGTGGACATGCGGCTGTCCCMACGTTCTGCCCCGT  
TTTCCCTGGGCCAMCCTTGGCCGTGCCACTGTGCTGTGCAACCAAMCTTA  
ATTGCCCTTCTTGGGCCCCAACCACTTTTGGTCTTTTGTAMCCCCATTTAT  
GTTTGTGTGAATTTTTTTTAAAKGATTTCAATWAAAATTTGGGCCCATTT  
TTTAAAAAATAAAAAAAA

## BSG2 (HBGBP46)

AGACCCACTTCCAAACCCCGGACCGCAGCAGTGCAGATCCACTTGCAGCAGGTCGACTT  
 T H F Q P P G P Q Q C R S T S Q Q V T F  
 TTTCAGGGACCCCTCCCCAGCGCTGAGCTGTGACCTGTCCCGGATCTGCATGANT  
 S G T P S P A L S C X V L C R I L H X W  
 GGCCAGCAGCAAAGTCCACAAGTTACTCAGCAGGGTCCCGCAGAGCTGCAGNCAGGGCG  
 P D D K V H K L L S R V A E S C X Q G P  
 CGGCCCTGCTGCTGGTGCAGACGCTCTGCTGATGAGGAGAGAGGGTGGCGCAGCGCTGNG  
 A C C W R R S W M R R R G W R S A X X  
 ATGCAGTCACTGAACATGCTGTGGTGCAGACTGAAGCCAGAGCGGAGCTGCGCAGTTTTC  
 A V T E H A G A D \* R Q E R S X A S F Q  
 AGTGTGTGNGAGNINCAAGGTTTCCACCAGTTCGCAAGTGTTCATTTNGGGGGGTT  
 C L X E X Q G F P P V A S C C I X G G F  
 TTTCCTGATGACGNTTTNGGCCACCAAAATTGCCCGTTNAAAGNCCNGENAGTTTTTA  
 S W M X X X A T K L P V X R X X X F F N  
 ATTTTNGGGGTTTNTTCCCGCAGGTTTNAAGTTGGANCCCGTTTCNNAAAGTTCCTT  
 F X G G F X F P R F X V G X P V X K V P L  
 TGGGAGTTTA  
 G D L

3/14

## BSG3 (HBGDQ94) FIG. 3

AGCNGNCITTCACCATCCCGTTTTCAGCAAGAGTCGTTACCCAGACAGGIGTTATGGA  
S X L F T I P F S S K S R S P D R C Y G

AGCTGCTGAAGGTGTGIGAAAATTTCGGAGAAAGTATTCTTACCAATTCAGAGCCCTG  
S C \* R L \* K F G E V I S Y Q Q L A A L

GNAGGCAACCCCAAGCCGCGAGCAGTTCGGAGGAGCAATGAGAGCCAAATCTGTCOC  
X G N P K A A R A V G R S N E R Q S C P

CATCCTCATCCCGTCCACAGAGTGTCTTNCAGCAGCCGANCCGTGGCAACTTACTTC  
H P H P V P Q S V L X A A X P W A T Y F

CGGAGGACTGGCCGIGAAGGAATGGCTTCTGCCCCATCAAAGGCCACCGTTTGGGAAG  
R R T G R E R N G F W P M K G H R F G K

CCAAAGNTTTCGAGGAGCTTCAGGTTTTCGCAAGGCGCTGNTTCAAGGAGCCCGENGT  
P X F G G S F R F W Q G P G F K G A G X

ACCTTCGGGNNCCGNCITTTTTCGNCGAAATNNNTTTTNCATTAGGNIGNTTT  
T F G X X R L F G R K X X F X H \* X G F

TTTTINGCCNCAAAAAGNTTTAAATNNATTGTTTCGGGNTTTCGNAACCTTTTTT  
F X P Q K X F K X X W F G X L X X P F F

TAAGGAAGTIG  
\* G S

4/14

## FIG. 4

BSG4 (HBGDA42)

GAGGACCGCCGACACCAACAGCTGCGCTCACTGACTG3CCCTCATCCGAAACCTGTTTC  
G T A D H H Q L R S L T G L I R N L X R

GGAACGCTAGGAACAAGGACGAGATGTCACGAAGGTGGTGAGCCACCTGATCGAGAAGC  
N A R N K D E M S T K V V S H L I E K L

TGCGCGCAGCGTGGGTGAGAAAGTCCGCCCCCAGCCGAGGTGCTGGTCAACATCATAGCTIG  
P G S V G E K S P P A E V L V N I I A V

TGCTCAACAACCTGGTGGTGGCCAGCCCCCATCGNIGCCCCGAGACCCTGCTGTAATTGACG  
L N N L V V A S P I X A R D L L Y F D G

GACTCCGAAGCTCATCTTCATCAAGAAGAAGCGGACAGCCCCGACAGTTAGAAGTACT  
L R K L I F I K K K R D S P D S \* K X S

CCCGGCAGCATCCAGCCTCCTGNCNCAACCTGTGGCANTACAAAGTTCACCCGTGAT  
R A A S S L L X N L L A X Q Q V P P \* F

TTCCGGGGAAGGTTTTCGGAAGGAGGATTTCTGGGGCCCTTAGGTCAGCTTTINGAGG  
P G E G F S E G G F L G P \* V K L X E E

AGAGTGACGTGNCANGINCAGGCACAGATTANTTCAGTTTTHGGACCCAGCTTGNCG  
R \* R X P X X G T D X F S F W T Q L X X

NGAGGTATT  
R Y

5/14

FIG. 5

## BSG5 (HBGBP36)

AGCCAGCCCCAGCCTGCCGGAGCCCCACGGAGCTACCTCTACCTCCACTC  
CTGCCACCGCCGCCCTCGACCAACGCCCTNTCCCCATCCGTCCCC  
NCTCAGGNAACCAAGATTNCAAGAGGGGCCGTGGCGTGCCAGGGCGC  
TAGGGAAGCCGGGTGGGGTGAGGGTAGCCCTTGAGCCCTGTCCCTGCCG  
CTGTAAAGAGCAGCAGGNACCCTGGGCCAGTCCAGAGACCTGGGGGTGT  
GTTGGGGGTGGGTGTGAGTNCGTATGAAATGTGTGTTGCTGGGGGCAA  
TTGTGCCCTGGAATCATGGGCAGGTNGNCCGNTCCGNNANGNCCGGG  
NTTNAATTNTTCCGTGGAAGNTTTAAAGGGTTGAATTTANGGTAAAAACC  
TTNGGGAAGGGAAGNTTCCAGGCAAAAAAGGTTT

6/14

## FIG. 6

BSG6 (HBNBF70)

AGACTCCCCGCTGGGTACCCCCCTGTGGCTGGCTCTGGGTGAATCATGGATGCCCTGAA  
D S P A G Y P P V A G S A \* I M D A L N

CAGCTGCCCCACGGACTGGCTGTGGGATCCCCCAAGATGTGGNAAGACAGCCCTG  
S L A T D S P C G I P P K M W X D Q P \*

ACCCCTGCGCGGTGTCTGCGCGCCCCCATCCAAAGCGCGGCTGCCCTGGCCACATCTACCCGG  
P L A G V C R P I Q G R P A S P H L P G

CCGTGGAGTTCCTGGGCCCCCTGGGAAGCAGNGNAGGAGAACTCGGCTCCAGAATCC  
R G V P G A L R S X X X G E T R L Q N P

ATCCTGCTGGTTCCGNCACITGGCCCCCAAGCCGNIGGTGCCCTGCCATTCOCATCTGCAGC  
S C W F R P L G P S R W C L P F P S A A

ATCCAGTGACTGNATCCCTCCTTNCATTTGAGTGGCCGTGTGTTCCANICATTNAGGTTT  
S Q \* L X P S X H L S G R C S X H X G F

TAAGCAGTTCNGNITGAGGTGCAGCCAAAGNCAAATTAACGGGCCCATTTNIGAGAAAAA  
K E L X X E V Q P X Q I N G X H X E K K

GGAANCAGGGTTTAAAAATNCAATTGGCGCACTTGTGTTTANITCATGTGTTAANGNA  
E X R V L K X Q L G Q L W F X S W L X X

AAAAAGTTTGGG  
K S L

7/14

# FIG. 7

## BSG7 (HBGBF39)

GCAGAGCTGGACATCGAGGGCGAGCTGCGCGGCTCATGGCCAAGCCCG  
AGGCTCTGAAGTCACTGCAGGAGCGGCGGNGAGCAGGAGCTGCTGGAG  
CANTACGTGAGCACCGTGAAACNACCGCAGTGATCATCGTGGACTCGCTGGA  
CGAGACCGGCTCCGGGGAACAAGAGGAGGATCAGATGCTGCGGGACATG  
ATTNAGAAAGCTGGCCCTCCAGAGGAAGAAGTTCCAAGTTCCGTTTGTTC  
AAGATCTTGGTCACCAAAAGCAAAAGCAGCCCTTCCANTNGTTAGCC  
CATTAGGCCCTTGGGTTTTGGCCGNAACCTTGGGAATTCGGTTTTTGGA  
NTTTGGGGGGNCCATGNTTTTTGGCCCNACCCGGGAACCCGGTTTTT  
TTAATTNGGGGNCCTNGGTTTTTGGNCGNACCNNGAATNTTGGGTT  
TTTTNGNTTGGNNAAGGGNTTTT



8/14

## FIG. 8

## BSG8 (HBGBU61)

AGGTCATCGGCAGCCTTCCCTGTGTGCCAAGCCAGCNTTCGCTTCTGAAA  
ACCAACTCCAGCCGCTGCCAGTCGGGACTTGGTCGCCCGCGCTGCCAG  
AATGNTCCACTGNCNACCGNCCCCCTGCCCTGGTTTCCCTTCTGTTTAGT  
GGGACACAGGCCACCCAGCTTTGGGTGGTGTGCTGACGCTCCCAGGGGTGC  
CAGGAGCCACTGGGACANGGTGAGGNTCCAGACGTTCTTCGAGGTGC  
CCAGTTNTCCAGGGAGNTTCTGGGCCCAAGCGTNTTNAAGGATCTTGTTTC  
CTTTAAACNNCCAATTG

## FIG. 9

## BSG9 (HBGCA96)

ACGAGAACAGANGATAGAGGGGCATCTNTCCCAGGTGACCNCTNCTTCTGT  
CCCAGGAGGTGGTAATTCCTNNGGGATGGGGCTCCCACACCTNCNTN  
AGGTCCCCTACTCAGACCCAGCACCAAGTGTCTGCCCTCTGAGAAATGTTGGCAGC  
TCACAGAGAGCAGGGCCGNCCTGATGGGGGCGAGTACTCCCACCTT  
CCTGCNTNCGATCCTANTTCTNATCCAGCGTCCNCTTTATTACCGTTTTTNA  
CTAATGCTTCTNTGAGNANGCTCTTTGGAAGNAGGAGCNNNAGCNTNNT  
GGAGCNTCNCANGAGATGTTTAAGGNTTATTAAAGCTT

9/14

## FIG. 10

BSG10 (HBGDM44)

ACATTTGGTGATGTNCTTGGCTGAGGTCTACTTGGCTTGCCTGCTTTGATCC  
TGAGAGNACCCACCCCATCTCACAGTGAATAGGTGGCAGGTGTGGGCTG  
GCGGTGGACTACACCTGNAGCTCCAGCTGCTGCAATCCGTGGAAGAG  
NCCATGGTGCACCTGCGGCTGGAGTGGCAGCTGCCCCAGGGACCCCAAGN  
CCAGCCTGTTGCCATGTCCCTCTCAGCAGACAATCCAGGTNCTCCTGGCA  
GAAGTGAAGNAGNCCAGACCCCTGATGAGCTCCTGGGTGAGGAGAAAGG  
TGTTCCAGGCTNTTTGGAGCCGCTTGGCCGATGGAGTTAAGGCCTTTT  
GAATTGTTTTGGGAGGNAACCTGGNTTTAAGGTNNTNAAGNCCTTGGCC  
CGNATTTTGGCTTCCAAATTNCCANTTGNNAATTTTTTTTNGAATTTGNTT  
AAGGTTGGGACTTTT

## FIG. 11

BSG11 (HBGBC21)

AGCATTGTGGGTAAAGGCCTGGAGGCAGGAAAGTGAAGGACAATTTCAAGAA  
ACTCNGTTCATCAATTTTCATCAACACCTTCCTGGGCCATGCCTGGTACTG  
AGGAACCCAGCCCTGAATCTGGACATCATTTTCCCTTTCAGAGCATAGAATG  
CAGGGGATCCAGGAATGGTTAACAGGAGNGAAGCTTGGTTCAAGGAGAC  
CTTTGCGTACCAGGNNAAAGGAGTTTGAACTTTANTCTNCANGGCAGGCA  
GAGCAGACAGGNTTCCTACCTGGGTCAATTCANTTNTCTATTCTCTTA  
NTGTTGCCATCCGTAGNTCCNCGGNTAANTGGGAGGTTTANGGCAGGNN  
GAATNGGTTTTGAACCCCNNGNGTTGGGGTTNGCAATTGNNGNTNNGNT  
GNGAACCATTTGAATNNGNGGCTTNGGTGNTCCAGGTGGGATCCCNNTTNN  
AAAAAAAAAAAAAATT

10/14

## FIG. 12

## BSG12 (HBGDN07)

GAGACGACTCTCTAGCCCGGCAGATGAAGGCGGANGGCGGCGCGNACT  
TGGATGAGATTACCGGAGCTGCGNCTCAGGGATCCGACGACCCAG  
CCCGACCCAAACGCCGAGTTCGACCCCGACCTGCCAGGGCGGCGCTGC  
ACCGCTGTCTGNNCTGCGGAGGTACTTCATCGATTCCACCAACCTGAAGA  
CCCACTTCCGATCCAAAGACCAAGAAAGGCTGAAGCAGCTTGAGCGTC  
GAGCCCTTACAGTCAGGAAGAGGCGGAGAGGCGGCGGTTATTGGGATT  
CTATGTGNCCCCCAGGNGNTTGGCAGTGCCACGGAAGTTTCCATTGA  
GGTNCCTNAGATTGNTAACTTTTANCTGAAATGGGCTTNAAGTTNCAAGG  
NAAAAGGANTTNCCTGGGCAATTACGAAAGNTTATGTTNGNAGGGAGGT  
TTCAACCCCTTTTNCCTTTGGTTTGGG

## FIG. 13

## BSG13 (HBGBD14)

AGGCACAGCTGTTGCGTCAGGCAAGGTCACCTGCATTATTATTGAGCAGC  
AGTNCCTGNTCAGGCCCGAGGGCGGAGCCCTCTCCCTGTTCCCTATGGT  
GTCTCCGAGGCCCTCTGGAGGGCCNCACATCTGGGAGCAGCACCTCAGA  
GTGGNACAGAAAGCATTAGCGTGCCACGAGCTTACCCGACGCCGAGCCT  
GTNAAGGTGGGCTGATGGTGCCCGTTTAAACCAGCGCTTCAGGAGGTTTC  
AGATGGNAGCCGAAACCCAGGGNTGTTNAGCATCANCTTCTGGAGNCC  
TTTTNTACTTTTATGACTGGTTTCTGGGAANGGTTGGTTGGGGAAGNN  
CANGNAGGCTTGGNGTTCCTTTAGGGTTNGGGGGCCTGTTTNNCATTENA  
ACCCAAAGCTTTNNNTTNCATTNNCTNTTTTTTNGGNGGCNAGTAAANN  
GNTCCAAGGTTTTTTT

11/14

## FIG. 14

## BSG14 (HBGBF11)

GGGCCACAGCTGCTCCTGCGCCTGGGCCCTTACTTCCTACCGAGACTTCCTG  
GGCACCAGCTGGTCCAGCTCAGCTGCCTGGCTGNGACANAAAGNTGACCCAC  
CGACTGGGTGACACGACGAGCCTATCTGGCGGACCCACTGGGGTGGGGC  
GCTGCACTAGCCACAGCCGATGGACTTCCTTGTTTCTCGCCGCTCCCCG  
GCAGGTGGGCTGAGGCCCTGGGCTGGTGGGACGTAACTGGTGGGGCAAC  
CTGAGCCTCAGGTGGAACCTGNNCGTTGTTTAACTGNAGGCAAGNCCT  
TGTTTNTGGGAATTGNCNNAATNGNGGACCAATGNTTGGNCNNAAGCC  
ANTTGNCCGATTTTTATNTNCCAAATTGAAAGTTNGNATTTTTGGNGGAAA  
NTTNGNGGAAGGAATTTAACTTNGANNGGGGGGNACCCGANGGGCCAAA  
GGGGTTTTAAANGAATTTTTTG

## FIG. 15

## BSG15 (HBNAC96)

AGCCACTCTGAGCTCCACGAGAAACACTGCTTCTCCAGGCCCGGGTTGT  
TGGGAGAGAGGCAGAGCAGCTGGAGCGCCGTNTCTCTCCTGCTGGGA  
CAACGTTTGGGCTTTGGGTATTGACTGAGTGGCTGACAGTTATCTTTGCAACC  
CCAAACTGGCTTTGGGGCCAGGACAAGGGGTNGGCNTTTATGGTGGTCCAA  
GTTTNNNTNCTTNCNACTNGGNTTGNCTCCTGACTGTGGANCNTGTTA  
ATTGGCTNTTTCANTGGGTTTTATTTT

12/14

## FIG. 16

## BSG16 (HBGDH36)

CCCCGTTGCCCCCTGCAAGCGCGCCCGCTTCGAGGACTTCTGCCCGGAC  
TCGTCCCCGGAGCGTCCAACATCTCAAACTTGATCTCCATCTTTGGCTCCGG  
CTTCTCGGGGCTGGTGAGCCGNACAGCCGACTCCTCGAGCAGCCGCCG  
CCGCTCAACGGGCAGCTGTGCGCCAAGCAGGCGTNCGCCAGCCTCGGCG  
CCTGGNACTCGAGCCATTGTNGCCTTCTAGGGACCCCGAGGGGCACAGG  
GGACCNGGGGCCCCGNGGGGTTGGGGCCAGACAAAGATTTGGNAAAGG  
GGCGAGAGGAGGGAACGAGNCGGGNCGGNCAATTGGGGTTTAAATTG  
GNNNNAANGGGNNNAANNNTTTTTTTTAAATNTTAAAAAATAAAAAA  
AAAAATTTGGGGGGG

## FIG. 17

## BSG17 (HBGDD46)

AGTNTCATACCGCCCAACCCGCTGGCTGGGAGGAGTCGGAGNTGAGAAC  
CTACACAGAGGTGGTGACAGAGTTTGGGACCGAGGTGGAGCCCGANTTTGG  
GACCAAGGTGGAGCCCNNGTTTGAGACCCAGTTGGAGCCTGANTTTGAGAC  
CCAGCTGGAACCCGAGTTTNAAGGAAGAGGAGGAGGAGGAGAAAGAGGAGG  
AGATAGCCACTGGCCAGGCATTCCCTTCACAAACAGTAGAGACCTACACAG  
TGAACTTTTGGGGACTTTTGGAGATCANNGTCCTACCAGACCCCGAGNCCAAN  
TTNAGGTTTNAAGCAGCAGGATTT

## FIG. 18

## BSG18 (HBNAI20)

AGGGCAGAAGCCATGGCCCATGNTGCACATCCAGGGAGGACAGAAG  
GCCCAGCTCACATGTGAGTCCTGGCAGAAGCCATGGCCCATGNTGCACAT  
CCAGGGAGGACAGAAAGGCCAGCTCACATGTGAGTNCCTGGNCAGAA  
GCCATGGCCCATGTTTGACATCCAGGGAGGACAGAAGGCCAGCTCA  
CATGTGAGTTCCTGGCAGAAGCCATGGCCCATGTTTGACATCCAGGGAGG  
AGGACAGAAGGCCAGNTCAGTGGCCCTGGCCCGNCAANTTANGNCCG  
ANCAGNAGGGTAGNTAGNCTTNTTGGNGTGTGTTCCATTCCGTATTNNG  
GTNNGGTTTTGGGNNCCANTTCAANAATTGGTTGGGNTTNTTTNNNT  
TTAATTAAAGTTTTAACTTNNAAAAAAAATTNNGGGGNTTNGNAT  
TTTTTTTNNTTTTGGGGGGTTTTAAATTNNCAAGTGAAATGTNTTAAATTTT  
G

## FIG. 19

## BSG19 (HBNAI67)

AGNGCCCTCCTGAGCAAGATCCACCTGTACACACGGCTGCCACAGCG  
ACCAGAGCCTTAGCCATCTGNTGTCACTGAAGCAGAGATGCTCAGGGACC  
CAGAGGTAGGCCAGCAACTGCTGCGGGACTCTGGAGCGTGAGAACCCAGCG  
CCTGGAGGCTGTCTGGCGTGGCGGCTCTGAGCTGGTCTTCTGGCGGAN  
GCAGCGNGCGCCTNGGAGCAAGGCTTGAGGCTTTGACGGGGGCCAA  
ANTTGAGTTCGCGCGGNGGCCCTTTTNGGGAAGTTTTGAAAANGAGCTT  
GGAGNTNTACAANATTTTTTNGNAGCNAAAAGTTGGCCCGNCCANCCCAT  
NGGCAAAAACGTTTGTTTANAAGGNAAGNTTTGGGNAAAAAGGACCCGGA  
CTTTGGGAAATTTTGTATTNAAAACCTTTAGGGNCCANAGNNTTNTNNG  
GGTTTAAATNGGTTAAAGGCAATTTTGAA

14/14

# FIG. 20

## BSG20 (HBGDQ22)

CANGCNTNTCCANGACCNTGGACCTCNCCTGGCCAGNCTGGANACCNCA  
TCGTCCCTATCAGTGGAGTNGCNTTACANGGCACTNGCGTCGANNTACA  
TCGCCNGGNNNNNTATCCATNTNAACTGATGTGGNGCATCAAACCCGCAG  
CTGTCGCCGCTTATGGGCACCCGGGCAGGCATCGCCAGGGAAGCTGGAGC  
TNGTGGAGCAGCAGTNTCGGCTGGAGCAGCTGAGTCGGCAGAGCTGCAG  
AGCAGNAACCAAGGCCACTGGGCTGACTGNCTACAGCGTACAGAGCCCCGG  
TTGGACAAGGACCTGGAA

# INTERNATIONAL SEARCH REPORT

International application No.  
PCT/US95/08295

## A. CLASSIFICATION OF SUBJECT MATTER

IPC(6) : Please See Extra Sheet.

US CL : Please See Extra Sheet.

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : Please See Extra Sheet.

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

Please See Extra Sheet.

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	WO, A, 93/16178 (VENTER ET AL) 19 August 1993, see entire document.	1-9
Y	Genomics, Volume 13, Number 3, issued March 1992, GIESER ET AL., "Expressed Sequence Tags and Chromosomal Localization of cDNA Clones from a Subtracted Retinal Pigment Epithelium Library", pages 873-876, see entire document.	1-9
Y	Cancer Research, Volume 54, Number 17, issued 01 September 1994, WATSON ET AL., "Isolation of Differentially Expressed Sequence Tags from Human Breast Cancer", pages 4598-4602, see entire document.	1-9

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:	*T	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
*A" document defining the general state of the art which is not considered to be of particular relevance	*X"	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
*E" earlier document published on or after the international filing date	*Y"	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
*L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	*Z"	document member of the same patent family
*O" document referring to an oral disclosure, use, exhibition or other means		
*P" document published prior to the international filing date but later than the priority date claimed		

Date of the actual completion of the international search

28 DECEMBER 1995

Date of mailing of the international search report

30 JAN 1996

Name and mailing address of the ISA/US  
Commissioner of Patents and Trademarks  
Box PCT  
Washington, D.C. 20231

Authorized officer

Lisa J. Hobbs, Ph.D.



# INTERNATIONAL SEARCH REPORT

International application No.  
PCT/US95/08295

## Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This international report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:  
because they relate to subject matter not required to be searched by this Authority, namely:
2. ☐ Claims Nos.:  
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:
3. ☐ Claims Nos.:  
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

## Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

Please See Extra Sheet.

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. ☒ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:  
1-9 and 14
4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest.  
☐ No protest accompanied the payment of additional search fees.

## INTERNATIONAL SEARCH REPORT

International application No.

PCT/US95/08295

### A. CLASSIFICATION OF SUBJECT MATTER:

IPC (6):

C07H 21/02; C07K 1/00; C12N 15/00, 1/20; C12P 21/06

### A. CLASSIFICATION OF SUBJECT MATTER:

US CL :

536/23.1; 530/350; 435/320.1, 252.3, 69.1

### B. FIELDS SEARCHED

Minimum documentation searched

Classification System: U.S.

536/23.1; 530/350; 435/320.1, 252.3, 69.1

### B. FIELDS SEARCHED

Electronic data bases consulted (Name of data base and where practicable terms used):

APS, STN(Biosis, Caba, Cancerlit, Cen, Drugu, Embase, Medline, Promt, Scisearch, Toxlit, Uspatfull), expressed sequence tag#, EST, ESTs, breast, human, breast specific, gene#, DNA#, cDNA#, clone#, subtraction clon###, cancer#, protein#, Seq. ID Nos. 1-28 searched on Genbank, NGenseq, EMBL, AGenseq, PIR, Swissprot

### BOX II. OBSERVATIONS WHERE UNITY OF INVENTION WAS LACKING

This ISA found multiple inventions as follows:

I. Claims 1-9 and 14, drawn to a DNA molecule, the polypeptide encoded by the DNA molecule, a recombinant method for production of the polypeptide and a method of patient treatment comprising said polypeptide. This group shares the special technical feature of the polypeptide, which groups II-VI do not share.

II. Claims 10 and 19, drawn to an antibody against said polypeptide and a diagnostic process comprising said antibody. This group shares the special technical feature of the antibody, which groups I and III-VI do not share.

III. Claims 11 and 12, drawn to a compound and a method of patient treatment comprising said compound. This group shares the special technical feature of the compound, which groups I-II and IV-VI do not share.

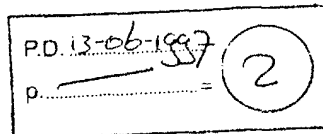
IV. Claim 13, drawn to a method of patient treatment comprising the DNA molecule. This group shares the special technical feature of patient treatment using the DNA molecule, which groups I-III and V-VI do not share.

V. Claims 15-18, drawn to a process of diagnosing a disorder by genetic means. This group shares the special technical feature of a diagnostic method, which groups I-IV and VI do not share.

VI. Claim 20, drawn to a method of identifying antagonists of the said polypeptide. This group shares the special technical feature of the antagonists, which groups I-V do not share.

Therefore, the claims are not so linked by a special technical feature within the meaning of PCT rule 13.2 so as to form a single inventive concept.

XP-002075 78



ID HS1255092 standard; RNA; EST; 337 BP.  
 XX  
 AC AA460323;  
 XX  
 NI g2185139  
 XX  
 DT 13-JUN-1997 (Rel. 52, Created)  
 DT 13-JUN-1997 (Rel. 52, Last updated, Version 1)  
 XX  
 DE zx51e06.s1 Soares testis NHT Homo sapiens cDNA clone 795778 3'  
 DE similar to TR:G207551 G207551 ODORANT-BINDING PROTEIN. ;.  
 XX  
 KW EST.  
 XX  
 OS Homo sapiens (human)  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria; Primates;  
 OC Catarrhini; Hominidae; Homo.  
 XX  
 RN [1]  
 RP 1-337  
 RA Hillier L., Allen M., Bowles L., Dubuque T., Geisel G., Jost S.,  
 RA Kucaba T., Lacy M., Le N., Lennon G., Marra M., Martin J., Moore B.,  
 RA Schellenberg K., Steptoe M., Tan F., Theising B., White Y., Wylie T.,  
 RA Waterston R., Wilson R.;  
 RT "WashU-Merck EST Project 1997";  
 RL Unpublished.  
 XX  
 CC Contact: Wilson RK WashU-Merck EST Project Washington University  
 CC School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis,  
 CC MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email:  
 CC est@watson.wustl.edu This clone is available royalty-free through  
 CC LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for  
 CC further information. Possible reversed clone: similarity on wrong  
 CC strand Seq primer: -41m13 fwd. ET from Amersham High quality  
 CC sequence stop: 323.  
 XX  
 FH Key Location/Qualifiers  
 FH  
 FT source 1. .337  
 FT /organism="Homo sapiens"  
 FT /note="Vector: pT7T3D-Pac (Pharmacia) with a modified  
 FT polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA  
 FT was prepared from mRNA obtained from Clontech Laboratories,  
 FT  
 FT Inc., and primed with a Not I - oligo(dT) primer [5'  
 FT TGTTACCAATCTGAAGTGGGAGCGCCGCCCAATTTTTTTTTTTTTTTT 3'].  
 FT Double-stranded cDNA was ligated to Eco RI adaptors  
 FT (Pharmacia), digested with Not I and cloned into the Not I  
 FT and Eco RI sites of the modified pT7T3 vector. Library went  
 FT  
 FT through one round of normalization to Cot5, and was  
 FT constructed by Bento Soares and M. Fatima Bonaldo."  
 FT /sex="male"  
 FT /clone="795778"  
 FT /clone\_lib="Soares testis NHT"  
 FT /lab\_host="DH10B"  
 FT mRNA complement(<1. .>337)  
 XX  
 SQ Sequence 337 BP; 57 A; 80 C; 119 G; 81 T; 0 other;

Hs1255092 Length: 337 September 30, 1998 10:27 Type: N Check: 6433 ..

```
  1  GGGGGAGAAG GACTTTATTT GGAGTCAGGT GGGTGGGAGC AGGGAAGGGT
 51  CATGGCTGGA GGGTAGGTCC AGGTGGTCCG GGCTCTGTGT CTGGTGGTAG
101  GGTGGGCTCT GGAGGTGCAG ACCCGGGGGC TGCCTAGTGT TCGGGAACGC
151  AGCTTCCCGT CTGCAGGGGC GTGAAAATGT CCTCCTCCGA GAGTCCCTTG
201  CGCTGCACCA ATTTCTTAAA TTCTTCCAGG GCCTCCCCGGT TGGTATCAGA
251  ATTCCTACCC ACAAGCTTTC CCATGTGGAG CAGGCCCCCA TGGTGCTGGT
301  CTTTGCAGTA AAAGATGTAG TGGTCCCTCC TGGGCAG
```

# XP-0C 079079

ID HSAC2098 standard; DNA; HUM; 36612 BP.

XX

AC AC002098;

XX

NI g2133900

XX

DT 16-MAY-1997 (Rel. 51, Created)

DT 05-SEP-1997 (Rel. 52, Last updated, Version 3)

XX

DE Genomic sequence from Human 9q34, complete sequence.

XX

KW HTG.

XX

OS Homo sapiens (human)

OC Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria; Primates;

OC Catarrhini; Hominidae; Homo.

XX

RN [1]

RP 1-36612

RA Hawkins T.L., Birren B.W., Fasman K.H., Nussbaum C., Lander E.S.;

RT "Genomic sequence from Human 9q34";

RL Unpublished.

XX

RN [2]

RP 1-36612

RA Hawkins T.L., Reeve M.P., Christoffersen A., Birren B.W., Fasman K.H.,

RA Lander E.S., McKernan K., Munro C., Richardson P., Barna N., Chang A.,

RA Cooke P., Daly M.J., Forrest C., Fripp W.J., Gage D., Geraigery K.,

RA Hagos B., Jacotot L., Lane M., MacKenzie J., Marquis N., McDermott J.,

RA Moloney N., Morrow J., Nachman A., Naylor J., O'Connor T., Peterson K.,

RA Rollins G., Spencer J., Stilwell J., Stone C., Strickland C., Sydney K.,

RA Wilmer F., Zody M.;

RT ;

RL Submitted (13-MAY-1997) to the EMBL/GenBank/DDBJ databases.

RL Whitehead Institute/MIT Center for Genome Research, 320 Charles Street,

RL Cambridge, MA 02141, USA

XX

RN [3]

RP 1-36612

RA Hawkins T.L., Reeve M.P., Christoffersen A., Birren B.W., Fasman K.H.,

RA Lander E.S., McKernan K., Munro C., Richardson P., Barna N., Chang A.,

RA Cooke P., Daly M.J., Forrest C., Fripp W.J., Gage D., Geraigery K.,

RA Hagos B., Jacotot L., Lane M., MacKenzie J., Marquis N., McDermott J.,

RA Moloney N., Morrow J., Nachman A., Naylor J., O'Connor T., Peterson K.,

RA Rollins G., Spencer J., Stilwell J., Stone C., Strickland C., Sydney K.,

RA Wilmer F., Zody M.;

RT ;

RL Submitted (29-MAY-1997) to the EMBL/GenBank/DDBJ databases.

RL Whitehead Institute/MIT Center for Genome Research, 320 Charles Street,

RL Cambridge, MA 02141, USA

XX

RN [4]

RP 1-36612

RA Hawkins T.L., Birren B.W., Fasman K.H., Nussbaum C., Lander E.S.,

RA McKernan K., Munro C., Richardson P., Barna N., Chang A., Cooke P.,

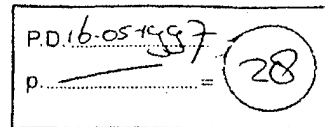
RA Daly M.J., Devon K., Dewar K., Forrest C., Gage D., Geraigery K.,

RA Hagos B., Harris K., Huang J., Hui L., Jacotot L., Kirby A., Lane M.,

RA MacKenzie J., Marquis N., McDermott J., Molla M., Morrow J., Nachman A.,

RA Naylor J., Nussbaum C., O'Connor T., Olotu A., Peterson K., Reeve M.P.,

RA Roberts D., Rollins G., Stilwell J., Stone C., Strickland C., Sydney K.,



RA Tang L., Wilmer F., Zemtseva I., Zody M. ;  
RT ;  
RL Submitted (02-SEP-1997) to the EMBL/GenBank/DDBJ databases.  
RL Whitehead Institute/MIT Center for Genome Research, 320 Charles Street,  
RL Cambridge, MA 02141, USA

XX  
CC The Staden databases, finishing information, and all  
CC chromatographic files used in the assembly of this clone are  
CC available from our anonymous ftp site.

XX  
CC All repeats were identified using RepeatMasker: Smit, A.F.A. &  
CC Green, P. (1996-1997)  
CC <http://ftp.genome.washington.edu/RM/RepeatMasker.html>.

XX  
FH Key Location/Qualifiers  
FH  
FT source 1. .36612  
FT /organism="Homo sapiens"  
FT /chromosome="9q34"  
FT /clone="161A1"  
FT /clone\_lib="Dr. David Kwiatkowski"  
FT repeat\_region complement(1. .85)  
FT /rpt\_family="AluSc"  
FT repeat\_region complement(243. .362)  
FT /rpt\_family="(CA)n"  
FT repeat\_region complement(1331. .1462)  
FT /rpt\_family="FLAM\_C"  
FT repeat\_region 8954. .9057  
FT /rpt\_family="MLT1B"  
FT repeat\_region 9066. .9860  
FT /rpt\_family="L1MD1"  
FT repeat\_region 9861. .10152  
FT /rpt\_family="AluSx"  
FT repeat\_region 10155. .10469  
FT /rpt\_family="AluSx"  
FT repeat\_region 10470. .11485  
FT /rpt\_family="L1MD1"  
FT repeat\_region complement(11500. .11591)  
FT /rpt\_family="LOR1"  
FT repeat\_region 12034. .12714  
FT /rpt\_family="L1MD1"  
FT repeat\_region complement(12730. .12896)  
FT /rpt\_family="AluSg/x"  
FT repeat\_region 12927. .14071  
FT /rpt\_family="L1MD1"  
FT repeat\_region 14067. .14166  
FT /rpt\_family="L1MD2"  
FT repeat\_region complement(15205. .15724)  
FT /rpt\_family="LINE2"  
FT repeat\_region 15832. .16155  
FT /rpt\_family="L1M4"  
FT repeat\_region 16246. .17496  
FT /rpt\_family="L1MB5"  
FT repeat\_region 17664. .17726  
FT /rpt\_family="MLT1D"  
FT repeat\_region 17730. .17795  
FT /rpt\_family="MLT1D"  
FT repeat\_region 17799. .17857  
FT /rpt\_family="MLT1D"  
FT repeat\_region 17868. .18260

```

FT      /rpt_family="MLT1D"
FT repeat_region 19360. .19656
FT      /rpt_family="MER67B"
FT repeat_region 19657. .19950
FT      /rpt_family="AluSx"
FT repeat_region 19951. .20271
FT      /rpt_family="MER67B"
FT repeat_region 20551. .20984
FT      /rpt_family="MER74"
FT repeat_region complement(21189. .21471)
FT      /rpt_family="AluJo"
FT repeat_region complement(21490. .21613)
FT      /rpt_family="(TA)n"
FT repeat_region complement(21614. .22056)
FT      /rpt_family="L1PA7"
FT repeat_region 26457. .26503
FT      /rpt_family="(CA)n"
FT repeat_region 27084. .27112
FT      /rpt_family="AT_rich"
FT repeat_region complement(27401. .27615)
FT      /rpt_family="(CATG)n"
FT repeat_region 27707. .28071
FT      /rpt_family="L1M4"
FT repeat_region 28072. .28358
FT      /rpt_family="AluSg"
FT repeat_region 29694. .30055
FT      /rpt_family="MER69B"
FT repeat_region 30080. .30143
FT      /rpt_family="AT_rich"
FT repeat_region complement(30146. .30417)
FT      /rpt_family="L1PA5"
FT repeat_region 30416. .30693
FT      /rpt_family="L1PA5"
FT repeat_region complement(34720. .34758)
FT      /rpt_family="(CAAAA)n"
FT repeat_region complement(35252. .35385)
FT      /rpt_family="AluJo"
FT repeat_region 36577. .36612
FT      /rpt_family="Alu"
XX
SQ      Sequence 36612 BP; 9620 A; 9585 C; 9441 G; 7966 T; 0 other;

```

Hsac2098 Length: 36612 September 30, 1998 10:19 Type: N Check: 1882 ..

```

1  GATCCTCCCA TGTCAGCATC CTGAGGAGCT GGGACTACAG GTGCGCGCCA
51 CCACGCCCCGA CTCCAGCATC TTTGTTCTCT TAATATTGAG TCCAGACATG
101 GAGTTAGGCA ACCGCAGGCA TGCAGATTAC ACTCCCCAGC AAGAAGGCTC
151 ATCATTGACA CACATGCAGA CTTAGGCTGT ACCAGGCATG TATACCAGGG
201 GGGTGCATTG TGAGTAGGTA GGTGTGTGTG CGGGTGGGAG ACTGTGTGTA
251 TGTGTGAGAG AGAGCATGTG GCAGTGTGTG TGTAGGTGTG TGTGCATGTG
301 TGAGAGCATG TATAGGTGTG TGTAGGGGTG TGTGTGTGCG CACATATGTG
351 TGTTCATGTG TGAGATTGTG TGTAGATGTG TGCCCGTTCC TCACAGCAAC

```

401 CCTGCTGGTG AGTGCTCCTG TGGTCCTCTC TTTACGAACA AAGAAACAGG  
451 TTAAGCACTT TCCGGAAGGC TGGTGGCCCA CGAGCCCTGT GTCACACTTC  
501 TCAGTCTGCC TGGGGCCCTA GCTCTCAGCC TGGACACTTG GCCCCTGGTG  
551 AACAAGTCAT GTCGGAGAGA ACCCGTCTGG TGCGCCATCT CCAGGATACT  
601 TGTGTCCCTT TGTACATATA TTTAAAAATT CATTTTGAAA GAACAAAAAA  
651 ATCTCCACTG CCCTCCTTCA CATCTGCGAT TTCTTGATTT TCAAACACTG  
701 AAAGGCTCGT TTTGCTCCTG GAGAGATTCC GCTCGGTGCC TCTGCCATGC  
751 TTGGGTTGGG AGTGAAAGTT TTCCTGACTT ATTTGGGATA CTGTCCTCAA  
801 ATATTTTACA GTTTGGGTGT CTGCGAGGGT CCCATGGTTA TGAGCCATGG  
851 GAAGAGGCTT CGTGTGACTC AGTGATTGGA GAGTTTGTGG GTGGAATTGG  
901 GGGGTGCTGA GCTTGGAAGG GGAAGCTGCA TGGGGCCAGG GCTCTCGGGA  
951 GCAGAAATGG ATGATGGTTG GGGCCACGT CCGCCAGGAC AGAGCCTCCC  
1001 ACCTCGAGGG CTGCATTCCC AGGGCTACAG GTCCCTGGTC CACCCATCAT  
1051 GAGGTGGCCA GGACATTGGA TGGATAGGGC AGATGAATGG ACAAACAGCC  
1101 CAGGCAAGGA TGCAGGGAGA CCCACAGCCG CCCATGCCAG GCCCTCCAAG  
1151 TGGGCAGGAA CAGGTTTCCC AGCTTCTTAT GTGCCCACCT CCTGTTCCCTG  
1201 CCAGCATCTC CCTGCACCCA CGCTCAGCGC TCTGCAGTCC AATCTCTGTC  
1251 TTTCCCAAGG AGACCCTCCC CACTGCTTAG ATGTGTTAAA AATCCTTTTC  
1301 CACCAGGTTA ATATTTTTAC TATTTACTTA TTTTTTCTTC TTTTTACAAA  
1351 CAGAGATGAG GTATCGCTGT GTTTCACAGG CTGTTCTCGA ACACCTGGCC  
1401 TCAAGTGTCC TTTCTCCTCG GCCTCCCAA ATGCTTACAG ATGTGAGCCA  
1451 CTGCACCTGG CCAATTTTCA ATATATGCAT TTTTTTTAAC TTTTATAGGT  
1501 TTCTATCCAA TAGCAAATAA CCAACCAATC TTGAATTTTT ACACATATCC  
1551 TAACCTTTCT GGAATCTGCA TGGTAGCACA GGGTGAGGGA GATGGCTTAC  
1601 TGCTCACCCA TCTTTGTAGG GTGAGGTCTG CCCTGCCCTC CACATGTGGC  
1651 CCGTGGACAC CCAGGTTGGT TCCGCCTCGG CTCGCCTGGG CCCTGCTGGC  
1701 CTGGCTCACT TCAGGACCAC CCTGTAGCCT CACCCAAGAA TCATGTCCAG  
1751 GTGGGAGAGG AGGAGATGAG GATGCTGGTC TGGCGGGAGG TGGGCAGTGG  
1801 TTACTTGGTG TGGCCGGGGG GAGAGGCGGT CGCTTTGAAG GCTCAGCCCC  
1851 CAGGAGGCCT GTCCAGCTTG GGTTTGTGGT TCCTGGGGCA TGTTTATCAC



1901	TGCGGTGCAG	TGGCCTCTCG	TCCCTCACTG	TGGTCGGCCT	CCTCATGTCC
1951	CAGCAGCCCC	AGTGTCTGCA	CAGACCATGG	CCCCTGGACA	AGCCATGGGA
2001	AGGCCAGGGT	GGGGCCAAGG	ATCTCGGAGA	CCAGGGTCGG	CTCTGCCTTG
2051	CTCTGGGGGT	GATGATGGAG	GCGCCAGTGT	CCCTCTCTCC	AGCGCCCCCTC
2101	TCTCCAGTGC	TGCCACTGAG	CCCAGGCTGC	TAAAATGGGG	GAGCTTGGAG
2151	GATCCCTCTG	AGACCCCTGA	GGCGGGAGCC	CTGCGGTGGC	CCCCTGCTG
2201	CAGCCCTGAC	GCCGTGTCCC	CGCTCCTGTC	CTCCCCAGTG	GCTTTCCAAA
2251	CGCTCCTGCC	CAGGGCCCTG	CCTGTCCTAA	GCTGAACCTG	AGACCTGAGG
2301	GCGGCTCTTC	CGGAGGCCCC	TCCAGGACAT	CACGCAGCCC	CTCAGGCCTG
2351	TTGGACATCG	TGCCTCTTGT	GGTTTGGCCC	TTGCTAGGTC	CAACATCTCC
2401	CCAGCTTCCC	TTGTGGCTCC	CCCTCCCAGC	AGACCTCCCA	GGGAGCGGGT
2451	GCAAGCCCCCT	CTGTCCCCAC	AGATAGGCCT	GCCCAGAGCT	GGGGGAGAAG
2501	GACTTTATTT	GGAGTCAGGT	GGGTGGGAGC	AGGGAAGGGT	CATGGCTGGA
2551	GGGTAGGTCC	AGGTGGTCCG	GGCTCTGTGT	CTGGTGGTAG	GGTGGGCTCT
2601	GGAGGTGCAG	ACCCGGGGGC	TGCTGTGCTG	GGAAGAGGAG	CAGAGGTCAG
2651	GGAGGCTGCA	GGGTAGGGCA	GAGGCCAGGG	CCAGAGCTCC	CTCCCCAGGC
2701	TCTGCAGCCC	CCACATCGGC	CACTGCTGGG	CAACCTCGGG	GCTCCAGAGC
2751	GACCTCAGCT	CCTCCAATGA	CTGGGCAGGC	CTCTGAGCAT	CCTTGGTGGG
2801	TGGATGGTGG	GGAGTGGACC	GCGTGGGGAC	AGGTGCAGGG	GGCGATGGGC
2851	TGAGCTGACA	CCTGAGGGAG	CCGGAGGGAG	GCCTCGCTGT	GGAAAGGCCG
2901	AGGAGGACCC	TCACTCGGGC	TGGCAGGTGC	AGGAATGATA	AAACAAAGGA
2951	CTTCCCGAAT	GTCCCAGGAA	CTGTGGGTGG	GCAGACTCTG	AGGTGCCAAC
3001	CTCGTGCCTG	ACCCTGGGGG	GGTCTCCCTG	GCTGTGTGGG	ATGGGAGGGG
3051	CCAGAGCAGA	TGTACCCTCG	AACCCCGGGA	ACCCAGGACT	CTGGGCTTGT
3101	CCAGTGCCCT	CCTAAAGGCT	CACTCACCCCT	AGTGTTCGGG	AACGCAGCTT
3151	CCTGCAGAGA	CCAAGAAAAA	CCCAGGGATT	AGAAGGCGCC	CTAGACCAGG
3201	GCCCAGACCC	CATCACAGCC	CAGAGCTCAG	AGCCCCAGAG	CCCATCGCAG
3251	CCCAGAGCGC	GGAGCCCCAG	ACCCCATCGC	AGCCCAGAGT	GCGGAGCCCC
3301	AGACCCCATC	GCAGCCCAGA	GCGCGGAGCC	CCAGACCCCA	TCGCAGCCCA
3351	GAGCACAGAG	CCTCCAATGA	GGAGGACGCT	AAACAGGGCC	CGGGAGCCCC

3401 TGCAGGAGAGA CTCGGGGCAC ATGGGGAATG CGGGGGACAT GGGAGGACAT  
3451 GGGGGTCATG GGACACTGGA GATAGCAGAC AGACACAGCA GAGGGACACA  
3501 GGGGACTGGG CACAGCCATC CTCACCCGTC TGCAGGGGCG TGAAAATGTC  
3551 CTCCTCCGAG AGTCCCTTGC GCTGCACCAA TTTCTTAAAT TCTTCCAGGG  
3601 CCTCCCGGTT GGTATCAGAA TTCCTACCTG CAGGTGAGGT GGCCAGGTGA  
3651 GCCGACGTGG GGACAGCGGC ACGGCCCTGC AGGGAGCAGA TGCCGAAAGG  
3701 AGACGACCAC GGCCCAGCAC CAGGACAGGG GGAGAGGCCT GAGAGTGGAT  
3751 CAGAGCTCGG GGGTGGGGCT GGGGACAGAG GAGATGGCCA CTGCCCAGCA  
3801 CCAGGACAAG GGGAGAGACC CAAGAGTGA CCAGAGCTCT GGGGTGGGGC  
3851 CGGGGACAGA AGAGATGGCC ATAGCCCAGC ACCGGACAAG CAGGAGAGGA  
3901 GACTTGAGAA TGGATGAGAG CTCAGGGGTG GGA CTGGGGA CAGAGGAGAC  
3951 AGTCACTGCC CAGATGAGAG GTCCCCAGGA AGGGGGACAG TGGCGGGGAC  
4001 TCTCCAAGCC ACTCGGCAAC CCCAGGGGAC CTGGGCAGCT CCCAATGCCA  
4051 CCCTGCAGGG TCACAGCCCT CACCTGGAGG GACCTTGTCT CCCAGAGGCA  
4101 CCGGCCACCC TCCTGCCCTG GAGGGTCCCT GCCCCATGAA GAGCTCTCTT  
4151 AGGACGCTTA AGGTGACCCT CCCCCTTAAG GTGACCCTCC TACACCCGAG  
4201 ACCCCAGAAG TGGCCCTTGG CAGGTTTCATG TCACTGAGGG CCATGTCTGT  
4251 CCCTGCCTGG CCCATCCCAT TCCCCTTCA GGTACTTGGA CCGGCTGCCC  
4301 AGCGGTGCCC GGCACATGAA AGCCGGCACA GGGGGCTTCC TTTTCCCCCA  
4351 TGCCAGGGCA TGGTGGTGCT GGTTCACCG GCTTCCAGAG GCAGAGACCG  
4401 TGGGGCAGGA CACGCAGACC CCAGCAAGCC CCTCACCAC AAGCTTTCCC  
4451 ATGTGGAGCA GGCCCCCATG GTGCTGGTCT TTGCAGTAAA AGATGTAGTG  
4501 GTCCCTCCTG GGCAGCTCCT GCAGGTACAT GAGCTTCCTG CCCCCATCTG  
4551 TAGATGACAG AGAAAATGGG TCATTCCCAG AGAGAACACT CGGGGCCACA  
4601 TGAAGAAACA CTCGGGAATG TTTCAGCGGT CACCCAGGTG CCCTGGACAG  
4651 CCTGAGCCAG GCCTGGCTGC TCCGCACAGT GTGGACCCGG ACACGGAGGC  
4701 AGGAGCCTCC TCTCAGGAAG GTCCCAATGC AAGTAATGGA GCCACAAGGC  
4751 CAGAGCACGT CCAGGACTAG AGTCCGGCCT TCCCAGCTCC AGCAGCTGCT  
4801 CCTGCCCCAC TTCCCTGTCC CTAACCCTCA TCTTCCCCCT CAGCCTCCCC  
4851 ATCTCTAATC CTCAGCTTCC GCAGCACTAA CCCTCGGCCT CCCCATCCCT

4901 AACCCCTCGGC CTCCTCATCC CTAACCCTCA GCTTCCCGAT CCCTAACCCCT  
4951 TGGCCCCCGT CCCTAACCCCT CAGCCTCCCC ATCCTTAACC CTCAGCTTCC  
5001 CCATCCTTAA CCCTCAGCTT CCCCATCCCT AACCCTCAGC CTCCCCGTAC  
5051 CTAATCCTTG GCCTCCTTGT CCCAATCCTC AGCTTCCTCA ATGTGTCAAT  
5101 GGCTAGGGCC TCCCAGAGGG CAGGCGTGAT CTGGTCCATC TCGACTGCGG  
5151 ACAAGCCTGT CACACCCGGT GTCTGATGTC AGGGCCACCA GCCCCCAGG  
5201 AAGGAGAGGC CAGCCCCCTCC TTGGAGGTGG GCAGAGCTGG GGCCCTGGCA  
5251 AAAGGGCAGC CTGGGCACTC TCTACCTGTG GAGGCTGGTG CACTGGGGTT  
5301 GGCGGTGGGG GAGGGTGGGG GTGGGAGTGG GGGAGGGGCT CACAGGCGCT  
5351 GTATTTGCCA GGCTCCTCCG TCTTCCGCAT CAGGATTTTC TTCTGGATGC  
5401 ACCGATCCTC CCTCCTGGAA AACAGGAGAC ACGCGGGCAG CGGCTCCCAG  
5451 GACACCCATG GCCCATCCTC AGCTCACCTG GTGCATGGCC CTGACCCGGC  
5501 AACCTGAAAA TTCCACTGTC CCCCACCCCA CCGAGCTTCA GGATGCCCAG  
5551 GCTGTTTCCC TCCAAGGCAG GGGGTGACTC TTCCCAACAC CCGAAACGTG  
5601 GATGGGGCAA TGGGCACCAG GTGGGTCTGG GGAGGGGAGC GAAAGTGGCC  
5651 TGAGGGGCCC TGCAGTGGGC AACACTCACA TGAAGGTGAA CGTGGCTTCC  
5701 AACTTCCCAC CGCCCAGGGC TGTCACCTTC ACTGGGGACA CCTTCCTGGG  
5751 CCTCCTGTCC TCCGGAAAGT CCTTATCGAC CACCATGGCC TTCACGTACC  
5801 AGGTCCCTGT GATCTGGAGC AGGCCAAGGC CGTGAGCCCA CCATGGGTGG  
5851 CCCAGATTCT ACTCTGACCC TGGCACTCAG GCCTGCAGCT ATAACCAGAC  
5901 ACCCATGGTG CCCGGCTGCT GCCCTTAAAG GCAGGCTGTG TTCCTGCACC  
5951 TTAGTTAGAG CTCAGCTGGA GTGAGTTAGA GCCAGCCCCC TGCTCAGGGC  
6001 TCCCAGCCCC CAGTGGAAGG AGGGCATGTC TGCACCCCAT GGACCCCCGG  
6051 GCCCCAGCAC CAGGTGAGCC CTCCCTCCAC AGGGCCCAAC CCTGCGAGAC  
6101 CTTCGGGCTT CAGGCGTCAG TGCAGCAGGA ACCCCTGAGA CGAGCCTGCT  
6151 CCGGCCCTGG GCCTCGGGGG GCCGTGTCTG CTGGAGGAAC AATAGGACCC  
6201 CTCCACCACC ACCCCAGGCT GGGAGCACGG GGTGAGAAGC CAGCCTTCAG  
6251 TGACACCTCC TAAAGCAGGG CCCCTGGCAC TGCCCCCTGC CCAGGAGTCC  
6301 GCCTGGCTCC TCCATCCGCC CACGCCAACC CAGCTCACAT CCTCCTCCTC  
6351 CAGGGTGAAG GACAGGGCAG CGGCCAGGCC GAGCGTGACA CCCAGGAACA

6401 GGGTCTTCAT CTCCAGAGCT CTGTGCTGCC GACCTCGGCA GGTCACCTGGG  
6451 CGTCTGAACA AGGCTGTGCT GGCTCCTCTC GAGATGTTCT TTATAGCCCC  
6501 CTGGCCAGTG TCCTGGGAGC ACGTGGCACG GTGCACCCCT CCCCATGGTG  
6551 GCAACCAGTC CTGCATCCGG GAGGGGGAGT GTCCTCATTG CTGGCATCTG  
6601 GTGAACAGCA AGTCAGTGCC AGCCAGAGAA GAAAAGACAC ACGATGCCAT  
6651 CCAGAGTTTG CACCACCCGA ACGCGCCCAC TCAGATGAGC CCAGAGGACT  
6701 CGAGAGGACA GGGCAGGGTG GTGTCCATCT GCACCCCAAG CCACAGAGGC  
6751 AGGGCCCACA GCCTCGGGAT ACTCCCTCAC CAACCCCTGC ACCCTCCCAG  
6801 GGCCACGGCT GGGACCCTGG GCAAGTCCTC TGGGGAAGTG GCCATCTCCC  
6851 TGCTGAGCAG AAATGGACTG GGCTTTAGAA CCAGAGAGAC GTGGGGCTGC  
6901 TGCCAGCTGA GTCCCGCACG CCCTAGAGCC CAAGCTTCCG CCCACTCTGG  
6951 GACCCACTCT GCATTGATTT TCCCCACCTG GACCCTGCAC TCGAGCTTCC  
7001 TAGGGGCAGA GGAGGGCACT ATTTTGTGGC TCAGGGTCAG TTTATATACG  
7051 GCCTCCTCGG GCCATATGAG GGACCCCGCC GCCTCCCCGA CCCCCTTCA  
7101 CTCCGTACCG GAGCCTGGCT GTGCTGTCCT CAGAACACGC GATCTCTCTG  
7151 GATTGTCCTG TCCATTTCTT CCTGCACGTA TTCATCGACC ATCTCCCTGT  
7201 TAGAATATAC ATTCCCCTCT CTCGTCAGCT TTGACGGCCC CCAGCCCCCA  
7251 GGACAGGCCT TGAATCTCAT CCCTTTAGAT AAACAAGTCG CATGAGGCAC  
7301 CAAGAGTCAC CTGCCAGCCT GTGCTCAGAG TTCGTGGCTG AGGTCCCTTG  
7351 GGCACCTGGC TCTGCTGGCC TCATCCCCAG CCAACCTGAG CCAGAACTTT  
7401 GCCTTCTCCC TCCCGGACCC TGCATTTCCC AGGTGTGTGT CATCTTCCCT  
7451 GGGACCCAAG GCTGACCCAC GCTCTGCAGG CAGGTGCTGG ACCCACACAA  
7501 TGGCATCTTT GTGCCTCTCT GGCCCAAGAT CAGACACTCT CAAGTACTTA  
7551 CTCAATAATA CTAATGCACA CTCTAATAAA CCCTGGGGCT GTAACGAACG  
7601 GGCTGGCGTG ACCCAGTCCC TCACCCCTCC CACGGCAAGT GCTGCCAAGC  
7651 AGAAACCCCC CACCCTCATC CCCCCGGGCA GGAAAGAATG GGGCCTGGGC  
7701 TCTTCAGGAC CAGAGATGAG CAAACAGGCT GGAGTCTCTA CATCAACGGG  
7751 AAGGTTTCTC TGTCTCTGT GTTTGTTGAC TGTACCCAG GCACCTCGTA  
7801 TTTCATTGGT TGTGTTGTGAA TCATTGATCT AAGCACCTTA CAACACTGGT  
7851 GCTGCTGGCT CAATCCTTGA AGAAAATGCT GGAACGACAT GCACAGATGC

7901 TTCAGAGAAC AGGGCTCTCC CATGAGGTCC ACGCTCAGTG CGACGGGTGC  
7951 TCTCCAGGAG TCTCCCATCA TCCACCCATC TTGCCCTGTG TCCCAGTCTC  
8001 ATTGCGCAAC TGTGTCCTTT TAGATACTGG TTGGGTTCAA CTGGTGCAGA  
8051 CACCGTCAGG AGATGGTAGG GCTGGAGGTT GGGGGTCAAC CCATGGCAGA  
8101 CATGAGGGGT GAGATCTCTG GCCCCAGTG CTTTGTGGCC TGGGGCACTT  
8151 CTCTGGATGG CTGTGTTTCT TCTGATGGTG GCTCCTGTTG GGCTGCCGCT  
8201 ACCACACAAC TTTCTCTCTC ACCACTTCCC TCACGACCCT TTTCTTTGAA  
8251 TCTCACAGTG ATCACAGCTT TTCAGGGCTG CTAATTCCTA GGTGCTTCAC  
8301 TGAAGTGGAC AGTGTCTGGG CTTTACTCGA CACCTGGGCA TAAGAAGGTT  
8351 TAGATGGCAC CACCAAACAA GCACCTAGAC CAGAGCATGT GCTGGCTGGG  
8401 ACTGAGGCAC GCCAAGAAGG AGCAGAGTGG GAGCTTACGA TGAATCTCAG  
8451 TTAAGGCTGC AAGACCAACT GCACAGTAGG GACCATCGCC TGTGCCACCA  
8501 ACCACCTATT TTAATTATTT CAGAGACGGT GGCTGGCCAC CATGCTGAAG  
8551 CATCAGTTAC AAAATGGACT TGATTTGCTG TGTGAGTGGG TCAGTGCAGT  
8601 GCAAGGGGTG GCTGAAGGCA CTGGTGACCC ACGCTGTCCG CCCACCTGCC  
8651 ACAGTTGTAG ATCGTTCCTG GTGCACGGAC GTATTACAG CTTCCCACCT  
8701 TAAGTGCCTG CACCTTTTCT TCTGTGCTTC CCCTGACACT GATGAAACCC  
8751 AGATGTGCAA GTGAATCAGC ATCCTCAGTC TGTAGGGATA GTAAGTATG  
8801 GGATAAATGG TCTGGACTCT TCCTTCTAAG TGGAACAATC AAATATGGTT  
8851 CCACCCTATT TCTCAGATGG CCCTCAGAAG GACTGAGTCC CAGTTGCCCA  
8901 CAATGGTAAC ATCATTAACA TATTTTGTG GTTTTTCTAC CTTCCAAATC  
8951 TCACTTGATC TCAGACTTCT GGCCACCAGA ATGTGAGAAA ATAAATTTTT  
9001 GTTCTAAGCC ACCCAGTTGG TAGTATTTTG TTATAGCAAC CCTATAAAAA  
9051 TAATACAAAC TCTAGCATGT TAATAATTAT ATTAAATGTA AATGGTTTAA  
9101 ATATTAAATG ACAGTGATTG GCAAAGTGAA TTAAAAGCCA TTATACAAC  
9151 ATATGCCGTC TACAAGAAAC TTAAGTCAAA TGTAACAATA GAGACAGGTT  
9201 GAAAGTAAAA ACATGAAAAA AGATATATAA ACATTAATCA AAAGAAAGCA  
9251 GAAGCATTAT TTCAATAACA GGTAAGATAG ACTTTAGAAC AAAGACTAAT  
9301 TAACAGAGAG GGAAATTATA CAATAATAAG TGGGCCATTC TGCAAAGAAG  
9351 ACATAGTAAT CCTAAACGTG TGTGCACCAA ACATTAGCTG CACAATATGT

9401 GAAGCAAAAC AGAACTGAAA GGAGAAATTT AAAAAGCCAC AATTATAGTT  
9451 GGAGACTTCA ATACCCTTCT CTCAGAGCAA CTAGACAGAA AACCAGCAGG  
9501 GACATCAAAG AACTCGGCCA CCCATCAATT AACATAATCG AATTGACATT  
9551 TGTTAGAACAC ACCACCTGGC AACAGCACTA CACACATTCT TTTCAAGTGC  
9601 ACCAGAATAT ATACCAAGAT AGACCATATC CTGGATGAGA AAATGAGCCT  
9651 CAAGAAATTT TTAAAAATTT AAATGATATA GAGGGTAGTC TCTGACCACA  
9701 TGGAATCTAA CTATAAATCA ATGCCAGAAA GATAATGGAA AATCTTTCTT  
9751 CAGATGCTTG GAAACAACAC ACCTGTGAAT AATCCACAGG TCAAAGAGGA  
9801 AGCCTGAAGG TAAATTTAAA AATATACTAG GCTGGATGAA AATGAAAATA  
9851 CAACATATCA GCTGGGCGTG ATGGCTTGTG CCTATAATTC CAGCACTTGT  
9901 GGAGGCCCAG GCGGGTGGAT CACTTGCGAT CAGGAATTCG AGACCAGCCT  
9951 GGCCAACATG GCGAAAACCC ATCTCTACCA AAAATGCAAA AACTAGCTGA  
10001 GTGTGGTGGT GCACGCCTGT AATCCCAGCT AATCGGGAGG CTGAGGCAGG  
10051 AGAATCACTT GAACCCGGGA GGCGGAGGTT GCAGTGGGCT ATCATGCCAC  
10101 TGCACTCCAG CCTGGGTGAC AGAGCAAGAC TCCATCTCAA AAAAAACAAA  
10151 AACAGGCTGG GCACAGTGGC TTATGCCTAT AATCTCAGTG CTTTGGAGCC  
10201 CCACGTGGGC AGATCACTTG AGGTCAGGAG TTCGAGACCA GCCTGGACAG  
10251 CATGACAAAA CCTCATCTCT ACTAAAAATA CAAAAATAAA TAAATAAATA  
10301 AATTGCTGGG CATGGTGGCG TATACCTGTA ATCCCAGCTA CTTAGGGGGC  
10351 TGAAGCAGGA GAATTACTTC AACCCAGGAG GTGGAGGTTG CAGTGAGCCA  
10401 AGATAGTGCC ACTGCACTCC AGCCTGGGAT ACAGAGCGAG ACTCTACCTC  
10451 AAAAACAAAC AAACAAAAAC ATATCAAAAT TTGTGAGACA CAGCTAAATC  
10501 AGTGCTGAGA GGCATTAAAT GCATATATTA GAAAAGAGAA GTCCAAAATC  
10551 AACTATCTAA GCTCCCCTT CAAGAACCTA GAAAAAGAAG AGCAAAATAA  
10601 ACCCAAAGCA AGAAGAAAAG AAATAACAAA GATAAGAGCA GCAATTAATA  
10651 AATCAAAAAGC AGAAAGCAGA AAAAATCGAA AGCAGAAAAA CAACAGAGAA  
10701 AAACAATGAA ACAAGGGTTA GTTCTCTGGA AAAATTAATA AAATTGAGGC  
10751 AGTTCTAGAA AGACTGACAA ACAAAAAAAG AAAGAAAGAC ACAAATAACC  
10801 AACATCTGGA ACAAACAGA GACTATCACT ACAGACCCTG CAGACATCAA  
10851 AGGATAATAA GGGAATACTA CGAAGAGACC TTCCACATA TATTTGACAA

10901 CATAGGTGAA ATGAACCAAT TTCTCAAAAA ACACAATGCC AATATAAGAT  
10951 AGATAACTTT AATAGCTCTA TGACAATTAA AGTTAACATT TTAATTTTAA  
11001 AAGCTTCCAA AAATAATTCT TCATGTCCAT ATGGTTTCAT TAGAAAATTC  
11051 TACCAACCAT TTAAAAATA AATAACACCA ACTCTGCATA ATCTCAGAAG  
11101 ATGAGAGAAT ATCTCCCAAC CAAAACCTGA CAACACAAAA AGGAAAGCTG  
11151 CAGAACAATA TACCTCATGA ATATAGATAC AGAAATCCTT TACAAAATAT  
11201 TAGCAAATAA TAATATATAA AATTAATTCT ACACCATGAC CAAAAGCAGT  
11251 TTATTCCAGG GATGGAAACC TGGTACAATA CTCAAAAATC AACCATATTA  
11301 ATATGTTAAG AAGAAAAATT ACATGATCAT ATCGATCAAT GCAGAAAAAG  
11351 TACTTGACTA AATTCATCAC GCATTGAGG AAACAACTCT CAGAAAAGTA  
11401 GGAATAGAAG AGAACCTCCT TAACTTGACA AAGAGCATCT AAAAAAACC  
11451 TATGGCTAAC ATCATTTCTC ATGGTGAAAG ACTGATGTGT AACTGGCTCA  
11501 AGTCCAGCTG CTCGCTGCTC AGAAGTCAAG GCATGAGAAG TGAAGTGTGG  
11551 TGAAAGGAAA GCAGCATTAT TCAAATGCTA GCAGTAGAGG AATGGCCAGG  
11601 CCCCATGCCT TTAAAAGACC ATTCAAACCT TCTGGACTGA GCGAAGGGGT  
11651 TTAAGGAGGA AAAGGTGTGG GAAATATGTG GGAATGGTGC AGGAGGATGT  
11701 TGGTCTGCAT CTTGTTCCAA TGGTTATCAT GAGTCATTAA TCGTCTGTCC  
11751 AGAGGTCTGG TTTGAGTCAT CCTGATTTC ACGGGAAGT GGTAGGCTAC  
11801 CTGTAACCTC CCCTAAGAGG GAGGATTCTG CAGCTGGGTC TCTCTGCCTG  
11851 GTTTGTTTCA AAATTGGCCC CTGGAATTC TAAGTAAGCA CATAATTAGA  
11901 TAAGCGAGCA CTGCTCACAG AAGTGCCTGG TGGGAAAGGG AGAAATAAAG  
11951 AGTTTCGAAG TATGTTTCAA GGCTGAAAGC AAGAAAGGAA AAAAGTTTTT  
12001 AAGCACATTT TGAGGCTGGG ATACTCAGTT ACAAACGCTT TCTCCTTAAC  
12051 ATCAGGGACA AGGCAAGGAT GTCCACTTTC ACCACTTGTA TTAAACATAG  
12101 TACAGAAAGT TCTAGTCAGA GCAATAAAGG TTTTTTAAAA AGGAAATTAA  
12151 AGACATACAA ATTGGAAAGG AAGGAAGAAA ACTGTCTGCA TCTGCAGATG  
12201 ACATGATTGT CTCTGTAGAA AATCCCAAGA AACTGAAAAA CAAAACAAAA  
12251 GAAAACAATA CACCTCCTAG AACTATTAAG CGAGTTCAGC AAGGTTGCAG  
12301 GATACAAGAT AAATATACAA AGATCAATTG TGTGTCTACA TACTAGTAAC  
12351 GGACCCTAAA CTAAAAACA CAATACTACT TATAGTGGCT AAAAAAAGA

12401	GATACCTAGG	CATAAGTTTA	ACAACCCATA	TGCTGGATTT	GTATGCTGAA
12451	AATAATGCAA	TGATTATAAA	ATAAATCAAA	GAAGATCTAA	ACAAATGGAG
12501	AGATGTGCTA	TATCCATGGA	TTAGGAGACT	CAACATAATA	AAAGCGTCAT
12551	TTCTCTCAAA	TAATATAAAG	ATTTAACACA	ATTTCTATCA	AAATTCAGAA
12601	AGGAATTTTT	GTAGACATAG	ACAAGATTAT	TCTAACATTT	ATATGGAAAG
12651	GCAAAGGAAC	TAGAATAACT	AAAACTTTTT	TTTTTAAGAA	TAAAGTGGGA
12701	GGAATCAGGC	TACCCTTTCT	TTCCTTTTCT	TCTTCTTTCT	TTTTTTTAGA
12751	GACAGAGTCT	TGCTCTGTCA	CCCAGGCTGG	AGTGCAGTAG	CACAATCTCG
12801	ACTCATTGCA	ACCTCCACCT	CCCAGATTCA	AGTGATTCTC	CTGCCTCCGC
12851	CTCCCAAGTA	GCTGGGATTA	CAGATGCACA	CCACCATAAC	CAGCTACAGG
12901	CTACCCTTTC	AAGACTCATT	ATATAGCTAG	CAGATGGACA	GATGCATAGA
12951	TCAATGGGAC	AGACAATAAA	GAATCCTCAA	ACAGACCTGC	ACGCATATAC
13001	CCAGTTAAGT	TTTTACAAAG	ATGCAAAAGC	AATTCAATGG	AAGATAGCTT
13051	TTGTAGCAAA	TGGGCTGGAT	TAAGTGGACA	TCCACAGACA	AAACCACCAA
13101	ACAAAAAGAA	GAACCACAAC	CTTGGTCTCA	CACCCTGCAA	GAAAATTAAC
13151	TCAAAATGGA	TCATAGACTT	AAACATAAAA	CTTAAACTA	TAAAAGCTTT
13201	AGAAAACAGA	AAGAAAATCT	TTGGGATATG	GAGCTGGGTC	AGGAGATTTT
13251	AGATGTGTCA	CCAAAAATAT	GATCCATAAA	AGGAAAAATT	GTAAAAATAG
13301	ACTTCATCAA	AATAAAAACC	TTTGTCTCTG	TAAGAGACCC	TGTTAACAGA
13351	ATAAAAAGAC	AAGCCAAAGA	CTGAAAGAAA	ATATTGACAA	ACTACATATT
13401	CAACAAAGAA	TTAGTTTCTA	TACTTTATAA	AGAATTTTTT	AAAACCTAAC
13451	AGTTTCAAAA	ATAGAATGTA	AACAAAAGAC	ATGAAGGGAC	ACTTGACTAG
13501	AGAGGATATA	CAGATGGCAA	ATAAACACAT	GAAAAGATGC	TCAACACCAT
13551	TAGTCATCAG	TGAAATGCAA	ATAAAAGCCA	TGATATATCA	GTACATACCT
13601	ATTAAGGTGG	CTGAGATAAG	ATCATAGTGA	TCATACTGAA	GCCTGGTGAG
13651	GATGCATGGA	AACTGGACAG	TTTCTGGAAA	ATGGCTTGAC	AGGTTTAAAA
13701	AAAAAAACCC	TGCAACAAGC	ATAGAACCCC	GCAATAGAAC	CCCTGGGCCT
13751	TTGTCCCAGA	AAAATGAAAA	CTGATGTCCA	TGCAAAAACC	TGTACATGTG
13801	TCATAGCCCA	AATCTATAAA	CACAGACATC	CTCCAACAGG	TGAATAGTTA
13851	AACAGACTGT	GGTACATCCA	TACCATAGAG	CACTACTCAG	AAGTAAAATG



13901 GAATAAACTA TTGATAAATG CAACAATGGG ATGAATCTCC AGAGAATGAT  
13951 GCTGCACAGG AAAAAAAAAA GCCAATCCCA CAATTTTATG ATTGCAATAG  
14001 ACGACATTCT TGCAATGAGC AGATTTTCAGG AGTGGAGAAC AGGTCTGTCA  
14051 TTGCTGGGGT TGGGGCTGGA GCAAGCCAGG TGAAGATACA CATGATAGCT  
14101 CTGTGTGGCT TTTTACAACCT GCATGGGAAT CTACAATGTC TCAGGGTTAA  
14151 ATGTTTAATT TAAAAACCAC AATAAGCATA AATGCTTAAA AAAAAAAAAAC  
14201 CTAGAGCCAA AGACTCAGCA ACTCCACTCA GAAGCATCAT TTCAACAGAT  
14251 ACACTCCCAC ACATGAACCA GCAGAAGGGT TCAATTGTTA CTTACAAATC  
14301 TTGGAAATGC CTCAAATTTT CATCAAAATA ACAATGACTG AATAAAGGTT  
14351 CATTTTTATG TTTGATTTTT GCAGACTTTC TAGATGCCTC TCATTCTTTG  
14401 GGGCACAAGG AATACATCGA TCAATGAAAC AAATGTCCTG TGCTCGGGGG  
14451 GCTTATATTC TAGGTGCAGG GGACTIONAAA TGAACACAAT AAATAAGTGA  
14501 ATTGCAGTGC CAGCTATCAA TGTCTTGCTT CCAGCTCCAC ACGTACCCTT  
14551 CATGGCCGGT TCCGAGGCAA TGGAGCTGGA GCAAGCAGGC GCTTCTCTTG  
14601 TGCACCTGGT GCAATGCTAA GCTTTGTTGG TTGAGGCGTA GGGAGAAGCG  
14651 CCTCCTAGCA TGGGTCCAGG ATGACTGGTT CCAGATGCTT CCATTTCTTC  
14701 CTTTTTCTTT CTCTTCTAT GCAAATGGCA GTGGCACAAG TCGGGGACGT  
14751 CTGGCGGCAC ACACCACATT GAATTCAGGG GACCGCCCCC AGATGGCTTC  
14801 CTATTGAGAT TCAGTGGCAT CTGCACCTGT GGGTTCTCAG TGACTIONTGC  
14851 AGGCTTCCCA GCCCCAGCCC ACCTGCCTTG GGTGGGGAGA GAGGCTTCCT  
14901 GCTTGCCAGA CCCAGCCTAC ATTTCCCTGC CTGCTGGCCT CTGCCCTGAC  
14951 CGTGGGCCAG CTGTGGCTGG GACAACCCAG CAAACTTCTC CAGCTTCCCT  
15001 TGGGAGAACC ACACCTTCTC CCATGAGGTC TGAACCCAGA CTTGTGGAAA  
15051 GGATCCCCAC CCCTGTTTAT TTCTTCCTTA ACTGCCCTCG TGCAATCCTG  
15101 AGCCATTCTT TGTGGTTCTC TTAAACCTC TAGAGCTGAT TGTTCATTGT  
15151 GCAAGCAATC CCTGTTCAAT CTATTGTCTC TGGATCCTGA TGGGTCACTG  
15201 GTTCATACGT TATTTTAAAT GGTGATCAGT GCTATGGAAA AAGCAAAGTG  
15251 GACAAGTTGG ACAGGAGTGT CAGGACAAGG GGTTATTTGT CCTTTAAATG  
15301 AATTGATGAG GAACAGCCCC GCCAAGGTGG CATTCAAACA CCGAATTTAA  
15351 GACATGAAGG AAGCAAGCCA TGAGCTCGTC AGCTGGAAGA GGAGTCCAGG

15401 TGGCAGGAGC ATCCAGTGCC AGCCTGGAGG TGGGAGAATG CTGGCAGTTT  
15451 CCGGGAACAC ACAGCAGAGC AGCCACAGCA AATGATCCAG GGAGACCGGC  
15501 CACACATGAG GCTGGGGAGT GGGTGGAAC TCCATGTAGG GATTAAGGAC  
15551 CACCACGAGG ACTTCAGCTT CTGCTCTGAG TGAAAGACCC TGGAGAGCTC  
15601 TGACAAAGGC AAGCTGTGAC TTACGTATTT AAGGGGTAC TCCAGTTGCT  
15651 GCATTGAGCA CAGATTTGGA GGCCGAATAT AGAAGCAGAG AAAGCAGTTA  
15701 GAAAAGACGC GGCAATAACC CAGGATGGTG CACTGATTTT TACCTGGGCA  
15751 TCCAGGAGAA CAGGGGATGG ACCTAACTGT CCCACCACAA ACAGAGGGAA  
15801 ACTGGCCCCC AGGTGCAAAA CGCTTGTGTT TAGAAACTGG ACACAGACAG  
15851 CGCAGGCTGT GATCCCTGAG AATGGGAGCT AACGAGCTCA GCTGGATCAT  
15901 GGCAGCTTTA TGCCTGGAGG GAATTTCAAT CCACCACAGA GAGAGGAACC  
15951 CAACAGAGCA AGGAGTCAGA CGGAGTGGAC GGGACACAGA TTGGACTCTA  
16001 GGCATCGCAG ACGACTAGAA TGTGGGAGAC AAGGAAAGAG ATAGGAGAGC  
16051 CCTCTGCAGA GGAAGAGCTC CAGAAACATG CAGAGTGGTC CTCAAGACCA  
16101 TAGCTAAATA CCAGTCCATG CAGACAGAGG GTGGAACTCC ACAGGGCCAG  
16151 AGAAAATGAT GTTTGGGTAA AGAAGGATTC CAAGGAACTG AAGAAAATCA  
16201 ACACCCACAA AACACAGGTG CAAGAATTAG TTGAGATGCC AAGATAAGAA  
16251 CTGTCTTCAA CAAATGGCTC TGGGACAACT GGATGTCCAC ATGCAAAAGA  
16301 CTGAGGTGGG ACCTCACGCC ACACACAAGC TTTCAACCAG AATGGATCAA  
16351 AGACCTCAAC GTAAGAGCTA AAAGTGTACA ACTCTAAGAA GATGTGTAAA  
16401 TCTCTGTGAC TTTGGCTTTG ACAATGAATT CTTCTCTATA CACCCAAATA  
16451 GCATGCAACC AAAGGAACAG ACAAATTGGA CTTTCATCAA ATTAATAACT  
16501 TTTGTACGTC AGTGGACACT GTCAAGGAAA TGAAAAGACA AGATACAGAA  
16551 CGTGAGAAAA TACTTGCAAA TCACATATCT GACAAGAGTC GAGTCTAGTA  
16601 TCAACAATAT ATAAAATACA CTTATAACTC AACATAAAAA CACAACTAGC  
16651 CCAATTTTAA AAAGGGAAAT GAATAGACTT TCTTCAGAGA AGATATAAAA  
16701 ATGACCAATA AACACCTGAA AATATGCTCA ATGTCATTAG TCATTAGGGA  
16751 AATGCAAATC AAAACTACAG CGAAATACCA CTTCACACTC ATTAGGATGG  
16801 CTAGAATCAC AAAGCCAGAT AATAACAAGG GTTGGCAAGG ATGTGGAGCA  
16851 ATTAGAACCC TCATGCACCA CTGGTGGGAA TGTAATAATGG CACAGCTGCT

16901 CAGAAAAACA TAGTTTGGCA GTTTATCCAA CAGTTACACA TAGAATTATT  
16951 TAATACCATA TGACCCAGCA GTTCCACTCC CAGGTATACA CCCAGGAGAA  
17001 ATGAAAACAT ATGTTCCCAT GAAAACATTT ATGCACAAAT GTCCCTAGCA  
17051 ACATTACAAA ACCCAAGAAG TAGAAACAAC TCAAATGTCT ACCAATGGAT  
17101 GAATGGATAA AAAAGGTGGT ATTATCCACA CAATGAAATA TTATTCTGTC  
17151 ATAAAAAGAA AATACCGACC CATGCTACCA CATGAGCAAA CGTTGAAACA  
17201 TTATGCTAAG GGAAAGAAAC CAGGCATAAG AAACACATAT ATATGATTCC  
17251 ATTTCTATGA AGTGTCCAGA ACAGGCAAGT TCATGCAGAC GGAAGTCGA  
17301 TGAGAGGTTT TGGGGCGTGG AGGGAAGAAT GGGGGTGAGA GCAGACGGGC  
17351 ATAGGGACCT TTGTAAGATG ATGAAGATCT CCTGGAATTA GATAGTGGTG  
17401 ATGGTTTCAC AGCCTTGTGA ATTTACTGAA ACCCACTGAA TTGGGCACTT  
17451 TTAAAGAGTG AGTATGATGG CATGTGGAAT ATATATCAAA TAAAAATTTT  
17501 AAAAACCACC TCCCCAGAGA AGCCTCTCCA GGCTACACAA TGGAAGATGC  
17551 AGCCCCTCCT CTCTCTGTCA CTTCCCTGAG CTGCGCTCCA TTAGGACACA  
17601 TCACTCCCCG GAGACACCCT GTTCATTTAC TTGCAGTCCT ATTGATTGCC  
17651 AGTTCCCCC ATGGGCAGAC TGAATAATGG TCCCCTGAAG ATGTCCGCAT  
17701 CCTCATCCCT GGCCCTGTGG ATGTGTGTAT GTGGTAGACT GAATAATGGT  
17751 CCCCTGAAGA TGTCCACATC CTCATCCCTG GCCCTGTGGA TGTGTGTATG  
17801 TGGTAGACTG AATAATGGTC CCCTAAAGAT GTCCACATCC TCATCCCTGG  
17851 CCCTGTGTCC CCCATGTGTA GACGGAATAA TGGTCCCCTA AAGATGTCCA  
17901 CATCCTCATC CCTGGCCCTG TGGATGTGTG CATGGTAAAA CGGACTTCAC  
17951 AGATATAATT CAGTGAAGGA TCTGAGATGA ACAATTACCC CAGATTATCT  
18001 GGGTGGGCC ACTGTTATCA CAGGGGCCCT TATAGGAGAG AGGCTGGGGA  
18051 GCGAGGGAGG GAGAGAGAGG AGGGTGAGGA GAGAGAATAT TTATGAACAG  
18101 AAGTAGAAGA GGCTGTGCTG CTGGCTTTGA AGGTGGAGGA GGGGCCACAA  
18151 GCCAAGGCAT GTGGAGGCCT CTGGAAGCTG GAAATGGCAG GGACAGACTC  
18201 CTGCCCAAGA GCCTCCAGAA GGAACCAGCC CTGCCAACCT TCGTTTTAGC  
18251 CCCGTGAAAC ATACTGAAAA GAAGAGCTGG AATGAAGCAT CTGAAGTTGT  
18301 CTCAGCAAAA CCTGCATGCT CGGTTGTGGT TGTGGGTTTT TTGACCAACA  
18351 AATGAAAGCG TTCTAAGACC CCATGCACCA TCGGCTGAAG GGGAGCATGA

18401 ATATCTATTC CCAGTCCCAG CCTCACAACC AGAGGATGCG TTCCTCACG  
18451 GTTTTCTCCT GCAGCTCAAG GAGGGAAAGC ATTTGGAATT ATTTGACTTC  
18501 TCCCCCTCAC TCAGTGGAAC TTTGGAGTCA TTCTTCCCAC CCTTGTCAAC  
18551 CACAGGGCTT TCTCAGGTCT ACAGGCGCGT GTGCAGATTG CTTGGGTGTG  
18601 CACAGATGCA GGGAGGGCAT CTCGTCGATA GAAAATGGGA TACTCTTGGG  
18651 CACGGCATGA AGTGACACCA CAATTCCCTA AATTGTTTGC TGTGGTTGCT  
18701 CATTGGGACG TGGCACCCAC TGGAACCAAG ACTTTGGGAA GCCAACTGGC  
18751 TGCTGTGCTT GGCCACAGAC GTGGTAATGA CTTCAGGGAG AATGGGGCGG  
18801 GGA CTGCCTG ACTGATCTAG GGCTTTCAGA AAGGAAATGC CAAGGTCTGA  
18851 GCTTCAGTAA CAGGCAAGAC AATGGAGATG TTCCATCCTG GCCTTAGAAT  
18901 AATATTTTCAT TCTTGCAGCC ACAGAGCAGA AATGGCTGAA AATCAGACTC  
18951 CATATTTAAT AGTGTGGATT GCATGGTGAG TTGGATTCAA AAGTCGCCAA  
19001 GTGTTTTATG TGAAAGTTGG GTCAC TGCTT AAGGAGTGGG ACCATGACAC  
19051 ACAGAACTGG GACATTTATG TGATCTCAGA TGGAAGTGAG AATCTTGATC  
19101 CCCTCCCCAA TCTCCCTGAA CCTGCCGGGG CAGAATCACC CCCTCCTCCC  
19151 CTACCCAAATC CCCCTGAACC TGCCTGGGCA GAATCATCCC CTCCTCCTTT  
19201 GCGTGAGGAG ATGAGACTTC CCTTTCTCAG ACATCTTACC CCGAGGCAGT  
19251 TGTGTAGGGG TGGATTCTAG GGGCTTTACA CCAGGAGAAC CGCGCTCCAA  
19301 TGTGGACCCA AGCTGAGTGT GTCAGTGCGT ATGTGCTGAG TGTGTCAGTG  
19351 CGTATGTACT GAGCACGGCA GAGCTCTCAG CATGGATCGG CTCCGCCCAA  
19401 GCCCCTGCAC CAGGAGACAG GAGATTGCCA AGCTCTCATG GCATCCAACA  
19451 GCAGGAGGCG CATCCCTGGG GTGGCCCCAG CGCCCTAAAT GCTTGCCTGA  
19501 GAGCTCGGCA CTGCAGGTGA ATTTGCTGTG TGTCCCTGGCC AGCACCCAGC  
19551 TGCCCCCTTCT TAGAGCTTTT GCTAAAAAGG GCTTACACCT GTGACTCCTT  
19601 CCTCTGCCAC TTTGAGAAGT CTGTGTTTCT CCTACAGTGC AGAAGTGTCT  
19651 TTCTCAGGCT GGCTGCAGTG GCTCACATCC GTAATCCCAG CACTTTGGGA  
19701 GGCATAGGTG GGTGGATCAA TTGAGGTCAG GAGTTCAAGA TCAGCCTGGC  
19751 CAACATGGTG AAACCCTGTC TTTACTAAAA GTACAAAAAT TAGCCGGGCG  
19801 TGGAGGCACG TGCCTGTAAT CCCAGCTACT CGGGAGGCTG AGGCAGGAGA  
19851 ATTGCTTGAA CCCGGGAAGC AGAAGAGGTT GCAGTAAGCC GAGATCGTGC

19901	CACTGCGCTC	CAGCCTGTGT	GACAGAGCAA	GA CTCCAAAA	CAAAAAAAAA
19951	GTGTCATTCT	CAGGGACCTG	AGAGCCATTC	CTTAGAAATG	TGACCTTCAG
20001	GAAGAATGGG	CCTCCTTCTT	CCAGACTCTG	TTGGGGGACA	GAGTCCTCCT
20051	TCCATAACTG	CCACACAGCT	GGCCCCACTG	CACTGACCCT	GACCAGCACT
20101	CTGTAACTCT	TCACTGGAGC	CCCCACTCCC	TCCCGCACCC	CCAACCACTC
20151	TCCCTTCAGA	ACACCCCTTC	TCCCCTGCAC	AAAGTGGCAT	GGAACCAGTC
20201	TGTTCCCTGGC	TCTGAGAAGT	TGCTGAATAA	AACCTGTCCT	CACCGCTTTA
20251	ACTAGCATTC	AGCTTGGCTC	ACTGGTGACG	ACGTATCCAA	AATGCCGTAT
20301	TTCACACATT	GGCTTGAGCA	GGAGAGCAGC	TCTCAGATGG	CTTCCAGGAC
20351	TGGCTGAGCT	GGTGTTGAGG	CCTCATTCAC	AGGGGCTGGG	ACGCCAGGAT
20401	GGCCCCACAT	AACATGGAGA	AAGGACTCTG	TGCTGCAGGT	GATTGAAGGG
20451	TTCCCAGGGT	TTGCTTTAGG	CTGGTG GGGG	AAGACACAGC	CATAGGAATG
20501	ACTGTGGGAC	GGAGGCTTAT	TACACTTAGG	CCCCTAGAAA	CAGGAGGGGC
20551	GGGGCCTCAC	GGAAGCACCA	GGGCCTGTCT	GGAGGCACCG	GGCGGAGCCT
20601	GCACTGCGGT	TCCCACGTTT	CCCACGGGAA	GGAACGGAGG	AAGCAGGGTA
20651	AACAGGCTCA	GGGCGGGCTG	GTTTCAATAA	TGTCCCGGGC	CCTGGGGTGT
20701	GGGGGCTGGC	TGTCCCTAGC	TGTCTGGTCC	CTGCCCTGGG	GAGAGTGTGG
20751	GCTCTGGA CT	GGTTGGTTTG	CATATGGAAG	GCACACTTAC	AGATGGTCCC
20801	TGCCCTGGGG	AGGGTGTAGG	CTCTGGACTG	GTTACCGGTT	TGCATATGGA
20851	AGGCACACTT	ACAGACAAGT	CATCTGCTAT	TTCTAGAAAT	TAGCCAACTC
20901	CTGGAGAAGC	ACTTTCTCCA	GGGTTAGCAA	GGCCCCAGAT	GTCAAAGTGT
20951	CCAAATTCAG	AAAGTAAAGG	CCATGGTTCA	CACAGAGAGG	GTGTCGGAGA
21001	CAGGTCGCGC	AGTCATTTGA	AGAGCAAATT	ACTGTAAAGA	ATCCCCAGCT
21051	GGTAACAGGA	GAGGGGCTAC	CAACGGGTAA	AGAGGACTCT	CAAACCACAG
21101	AAATAACAGA	AATGGGAGCC	AGCTCTACCC	CAGGGCTGAG	GCTGCACACC
21151	GGGGTAGGAA	CAACCTGGAA	GGTGCTCAGG	CCTCTGTGGA	GAGGGTCTTG
21201	CTCTGTCCCC	CAGGCTGGAG	TGCAGTGACA	TGATCATAGC	TCATTGCAGC
21251	CTCGACCTCC	TGGGCTCAAG	CAATCCTCCC	ACTTCAGCCT	CCCAAGTATC
21301	TGGGATTACA	GGTGTGCACC	AGCAAGCCTG	GCTAATTTTT	GGGGTTTTTT
21351	TGGTAGAGCT	GGGGTCTCGC	TATGTTGCCC	AGGCTGGTCT	CCAACTCATA

21401 GCTTCAAGTG ATTCTCTCAC CTCAGCCTCC CAAAGTGCTG GGAACACAGG  
21451 CATAAGCCAC CATGCCTGGC CAGAATTGAG TTTTTTTTAA TATATATATA  
21501 CACATATATA TACACACACA CACACATACA CACATATATA CACACATATA  
21551 TGTATATATA TACACATACA CATATATATA TACATATATA CACATATATA  
21601 TATTTTTTATT ATACTTTAAG TTCTAGGGTA CATGTGCACA ACATGCAGGT  
21651 TTGTTACATA TGTATACATG TGCAATGTTG GTGTGCTGTA CCCATTAGCT  
21701 CGTCATTTAC ATGAGGTATT TATCCTAATG CTATCCCTCC CCCTACCCCC  
21751 CACTCCACGA CAGGCCCCGG TGTGTGATGT TCCCCTTCCT GTGTCCAAGT  
21801 GTTCTCATTG TTCAATTCCC ACCTATGAGT GAGAACATGC GGTGTTTGGT  
21851 TTTTGTCTCT TGTGATAGTT TGCTGAGAAT GATGGTTTCC AGCTTCATCC  
21901 ATGTCCCTAC AAAGGACATG AACTCATCAT TTTTATGGC TGCAGAGTAT  
21951 TCCATGGTGT ATCTGTGCCA CATTTTCTTA ATCCAGTCTA TCATTGTTGG  
22001 ACATTTGGGT TGGTTCCAAG TCTTTGCTAT TGTGAGTAGT GCCGCAATAA  
22051 ACATACCCCA GGGCTTCTTG AAGCGCCACT GAGCCCCAGG GCTGGATTTT  
22101 CTCAAATGTC CTCTAAAGTT GTTCGCTCAC CCTAATCGCT CCCTGGAGAG  
22151 CAGGTTTCTG CAGCAAAAGG AAGCACAGGC GAGTCAGGAG ACCCCAGGAA  
22201 TGCAGTCAGG ACCGGGGCCC CTCATCAGCC ACCAACTCCA CCACCCCAGG  
22251 AAGGACAAAG CCAGAATTAG ACACAGTACA GTTACGCGCC CGTGCAATGC  
22301 AGCTCTGCTG ACGTCCCCAT GCGGGGACAT GGGGGGCTCT GCAGGGCCAT  
22351 GCCTCCTACC GCTCTGCCTG GGGATGAGGA TGCTCTCCGT GCTGAGTCCA  
22401 CGGGCTCCTG CAGCTTTCTC AAAGTCCTCC AAGGCTTCCA GGTGTTCTC  
22451 GGGGTCTCTG CCTGTGGGTG CCAGGACGAG ATGCGGGTGC TCAAGGCAGC  
22501 ACTGGGCAGA CGGCACCCCG TCCCAGCCTG GAGGCCTGAG CTAGCCTGGG  
22551 GAGAGGGACC TTGCCCACGG TGGGGGGA CTGAACCCAGA GGTCTCAGG  
22601 CTGGCAAAGT GGGACCAGG TGCCGTGCAC GGCTGAGAAG GCTCCCGCAG  
22651 CAGCAGAACC GGGAGAACGG GAAGGGCTCT CCGTCCATCC CCAAGCATCT  
22701 CCCGGAGTGA GAATGGACTG CAGGACAGGC CCGCCCTGGA GGAAGCGCCT  
22751 CTCCCCTCGG GGGCACCAGG GCACCGCCTG TCCCATTCTCT CTGTCATTCA  
22801 CACACCCCTG GCTCCAGGGA GAACTCAGGA GCCTTCACGG GCCTCGTCCT  
22851 CCGACAGGGA CGGCACTGAT GCGCCACCGT CTGTGGGAGC CACCGCGTCT

22901 GGAATTCAAC CCCGGGAGCT GGAACGTTCT TTCCACGCAG ACTCCTGCTC  
22951 ACATCCCCAT TCCTCTCCCC TGCTGGACAC CCCACGGGAA CGCGCTCAAG  
23001 GATGACACTG CAGGGAGCTG CCCCAGGGGA GGGCCTGGGC CACTCTGGGA  
23051 CACGTGTGAC CTTTGAGCCC CTGTGCCGAC AGCTGTGGCT CCTTCCTGCC  
23101 TCTGTTGCAG AGACCGAGGC CGGGGCATCC CAAATCCTGA ACTCCAGGGT  
23151 GGGAATTGGG GCCTTTATCA GGAAACTCA TCCTAAGGAA ACTGCAGTGT  
23201 GTTCCTGGGT CCCAGTGGCT TCTCCCATGG GGAAGAGGGT GCAGGAAGCC  
23251 GAGGGAGGAG GGAGGGTGGG TGGAGGCATG GGTTCAGGG TGGGGGTACG  
23301 GGACCCACCC ACGAGCTTCG CCCCTCGGAT CGGCTTCCCG TGCAGCTCAC  
23351 CCTCACAGTA AAAGATGTAG TGGTCCTTCA CGTGCGACCT GATGATGTAT  
23401 GCCACGTGCT TGCCCCGTGC TGCAGAAGGA CAGCAGCATA GTCCTGAATT  
23451 AGCCACCGGG GCCTTCAGAG AGGATCCCGG GACAGTGTGC GTTGCCAGGC  
23501 CCTGCAAGCC CCGCAGCTCC TGCACCCTTC CAGCCTCCCA GAGCTCCCTG  
23551 ACCCAGGTTC CCCACTGGGC CTGAGCCCTA GGCCATGGGA CCCCCACTTC  
23601 CCAGGAACCT GATGCACCCG AGAGGTGGCC ACATCTCAAA CCAGTGAGTC  
23651 ACGGGGGGTT CCTCCGGGTG TGAGGTCTGC GGAGGTGGGA GTTGGGGTAC  
23701 AGCCCCCGC AGCCAAATAT GACTCTGCTC CTGGCCCTGC CAGAGCCGCA  
23751 GCTAGGGCTA CAGGGAAAGC CCAGAGGGTC CACCCAGGGA CAAACCCGT  
23801 ACACCTGGTG TTTGGGGTCA GGGGTTACA TTTCTAAAGC AAATTGACCA  
23851 GTGGGAGAGG AAAGTGGGCT TTTAGTGGAG TCAGGGGCAG CAGGCCAAAA  
23901 GGCCACCCCA AAAAGGCACT GGCTACCCCC CCACATCCAG CGTGTGTTTG  
23951 AGTTTAGGCT CTGGCTTAGG CTCCGGGACT CACTGGCCGT GTATTTTCCC  
24001 GGCTCGTCAG TTTTCTCCAG GACGACCTTC ACCTCCTGGC ACTGGCCACT  
24051 TATCCTGGAA AACAGAAGCC CCCTGGCAAA GCGGCCCTGA CCTGTGCCTC  
24101 CCCCAGCCCT GGAGCAGCCA CGCTACAATG CCCTGCAAAG GTGGCCCTGA  
24151 GTCTGCCACA GGCCCATGGA GGAGGATGCT TCTCTGCATC TCAGACCCAG  
24201 GCACCAGCAG CCCCTCCGGG TCACACATGG CAGTCCCGG GGTCTCCTGA  
24251 GGTGGCCAGG AGCTCTGCAT CGCTCCTCCT GGTGGCTGCC TTCCCAACCC  
24301 CCCCAGCCCC AACACATTTT CTGCCAGATG CACCCGGGCT GCCTGGTTTG  
24351 TAGGCAGTGT CTGTGGCGTG GGCCTGAGCC TCATCCCACC AGCTGCATTC

24401 TTCCTTTTAG AGACATTCAG GGACAGGCCC CATACCCATC ATCATCTGGC  
24451 CCAGAGAGGG AAAGGTTTCAG TGAGGGGCTT TTGGGGAAAA AAAAAATCAG  
24501 GGTCATTGCC AGGGTCAGTG ACTTCCTAAG GGTCAGCCCC CAGGACCGGA  
24551 GCACAGGCAC TGCATGCCAC TCAGGAGGGA GCCACGCGGG TGAAATCACA  
24601 CAGGTGGGAA GTCACAGATG TGAGCCACGC GGATGGAAGC AGCACAGGTG  
24651 AGAGCCGCAC AGGTGAGCGC TACACAGGGA CTTCAAGGGG GTCTTCTCCA  
24701 TGGTGTCTTA TGGCCCCCTCT TGGTCATTCT GCAAGCCCCA TCCCTGCGCT  
24751 TCTCCAAAGC AGAAATTTGC TGCCAAGACT ACAAGTCTAG AGTCTGTAAG  
24801 CCCTGCCTGC TGGCTCTGAC TCTGATCGGG GGTAAACACC TTTTCCCAAA  
24851 GGGGCTTGCA CGGCACCACA CATTCCTTAG TCGGAGGAGG TGGTGGGTGG  
24901 TACCCAGCAC TGCCCCTGAT AGCCACAGG GGCCTCTACT CCCTCAGGTC  
24951 CCCTCTAGCT GAGGCACCTC ACTGGCACAG TTTGCCCTCA AGCATTGCCT  
25001 GAGCAGGCAG GGCTGAGAGA GGCCTCCCAG CAGGTCTGCC CACCTCCCCA  
25051 GCACCAGAGC TCTCCTGGGT GGGGAAAGGG AAGGCCCTAG ACCAGGTTGC  
25101 AGGCTGCCCC GCCGGCTGGC AGACACTCAC AGCATGGTGG CCTTAGCTTC  
25151 CAGGTTGCCC CCTTCCAGGA TTGTGAGGGT CATGGGTGTC ACCGATTCCA  
25201 GATTCATCTC AGGGAGCTCC CTGTCCACCG TCATGGCCTT CAGATACCAC  
25251 GTCCCTGACA CCTGGAGAAA GTTCCTCTCT ATCAGGCTCA GGCTGACCGC  
25301 TCCAGAGCCT CCCCCTTCCC TGGCCTGCAG CCCCAGACT CTACCGCACC  
25351 CCAAGGGCTG GAGATGGGGA ACATTCCCCA ATCTAATCCA CCCAGAAAAC  
25401 ACCCAGCACC CTCGCCAGGT CCACCCGAAG GCTCCAGCCC CCGCAAGTTG  
25451 GCCAGGACCA CAGGAGCCCT CACCCCAGG AGAGAAGTGA AGTCAGCCAG  
25501 ACCCCCACCC TCCTAGCCCT GCCCTTCTGC ACCCCGGCCC CCTCCCAGCC  
25551 CAGACTCCTG CTGTCCCATC TCACCATCTT TCAGGGCAGG GCTTTAATGC  
25601 CCAACCTGAA AATGAGGGTC TCCTTGATG GGGGAGGGAT GTGGTCTCCA  
25651 TCCAGCAGTC TCAGCCTCGC CCCTTGCCCC CTCCAGCCCA GCCTTCCCAT  
25701 CCGGGCCTCA CATCCTGAAT CTCCTCGTCT GAGGCCAGGA GGTGGTGGGC  
25751 CTGCAGGGCA GCAATGAGGC TGAGGCTGAT GGCCAGGAGC AGGGGCTTCA  
25801 TCTCTGGTGT CTGAGTTCAC GGCCGCCTGA CAGTTCACTT GCTGGGGCTG  
25851 GGAGAGGGTG TGCCACTCCC ACAGCCGCCC TTTTACAGC TGGGCTCAGG



25901 ATCCCCTGGG GACCTAGCAC AGGTGACTGA CCAATCCCTT CCTGGATGTA  
25951 AACCACGGCC AGTTCTGCAA CGGAGCAGAT AACGTGCCAT TGTTGCTGGA  
26001 GGCTGGTGAA TAACAGCTTG TTAATTCAG AGGATCATCC GGACAGGACC  
26051 AGGTCCAGAG GAGCGGGTTT GCAGATTTCA CCTGGGAGGA AACTGAAGGG  
26101 AGTGAAGGGC GTGCAGGCCT CCTGGGGACA TGTGCGGCAG TAACAGGCGC  
26151 CCCCCAGCCA TCCCACATGC GCTGATGTGC ACACCGTCAA GGGGGTTTGG  
26201 GGTTGGAGTC CGAGGGCCGT GGGCCTGGGG CTCGGCTGAG CTTCCTGCAT  
26251 CTCTGGGCAG GTTACTTGGC CTCAGATTCA TCATGGAAAA AAAGGGACCC  
26301 TCCTGTGTGA ACACGATACC GTGTGCATTT AAGGACGGTG TCCACTGTCC  
26351 TGAGCAAGGT GTGCAGTAAC GACCCTCTCT GCTCCACACA CAGAGCAGCA  
26401 ATAGGTCAGA CAGACAGAAG ACCAAGACAG GTGTGTGTGT GTGTATGCAG  
26451 GTGTGTACAT ATATACATGC ACACACATGC ACAAACACAC ACACACTCAC  
26501 ATAAAACCAG GAATGGAGGA GAAGCACAGG CTGGTGATGG GGCAGAGGCA  
26551 GGGGCCTGTC AGTTCCAAGT CCAAGGCAG GCAGAGAGGG CCTAGATCCA  
26601 GCTCCTGGGA GTGATGGGAG GGGCTTCCCT GACTCCAGAA AGGAGGCTGA  
26651 GAAACTCTCC TTGACACATC CCAGGGCTGT AACTTCTTAT AACTTAGTGG  
26701 AGCACTGTGG GCATAGGAAG GCAAGAGGAC ATCGACACAG CCTCAAGGGC  
26751 AGAAGCTGAG ACAAGCCGTC CTCAGAAAGG GTGACTTTAT CTGTGACATC  
26801 CTCCACATCA CTGCAGAGCA GAAGCCTGGG ACCCGGGGCA GAGGCTACGA  
26851 AACACTGCTA GACAGAGGAC AGTGACAACA AGATGTCGGA GAAAGGCAGA  
26901 AATGCAGAGA CTCAATATCC TTGCCTTGAA ATAAGCTAGA AAGGCCAGGT  
26951 TCCAAATTAC CCAAGGCAAC ACCACACCAA GAAAGGCAGG CTAAAAATGA  
27001 GCAATTGGTA CAGGAACTTA CACCAGCTAA GGAAAGTCTG ACACACACTT  
27051 TAAAACTGCA GGAGTATGAT CAAGATGCCT AGCATTTTAA AAATATATAA  
27101 AGTATAAAAT TAGAACTGGC CAGAATATAA GAAACAAGGA GAGATGAAAT  
27151 AGAAATAATT ACAAATCTTG AAGATGAATA ACAAAGTGAA TAAAAAACT  
27201 CAATAGACAG GTTAAATTCC AGACTTGACA TAGTAGAAAA GAGACTTAGT  
27251 GTCATGGAAG ATAGTACCAA CAAAATCAAT GAAGACAGAA AAATTAGAAA  
27301 GAGAAATTGA GGTGTTACAG ATAGATTCAG TCTCCATCCT GGAGCTAGTA  
27351 ATAAAATGAG TGGAGGAAAC CATCCAGAGT GCAGCAGCTC TGAAAACACA

27401 ACACACACAG ATATACACAC ACACATGCAC ACACACATGC ACACACACGC  
27451 ACATGTGCAC ACATACAAAT GCACACCCCC CATAACATG CACACATACA  
27501 TGCACACACG CAGACATACA CACGTGCACA CACACCATGC ATGCATGCAT  
27551 ACATACATGC ACATACATGC ACACATACGT ACACACATGC ACATTACAC  
27601 ACATACACAC ATGCACAAAG TTAACACACG TGGAGAATAG ATCAGGAGGC  
27651 TCCAATATAA ACAGAATAGG AGTTCCGAAA GAAAAACATT TTCTAAAATG  
27701 GCAACAAAGC AACGGAATAG CTCAGAATTT TTCAGAATTG AAAAGACATG  
27751 AGATCCCAGA TCTAAATTAC CAAGGAATGA TAAAAATTGA GAAATCCACT  
27801 CCTAGAATAT TACAGTGAAG TGCAGATCAC CCAGGAAAGG AGAAAAAAG  
27851 TCAAAACGAC CAGAAAGAAG AGCCACACTA CATACCATAA AAGCGGATGT  
27901 GGACTGATGG CAGACTTCTC AATAGCAACG CTGCCTTCCG GAAGACAGTG  
27951 GAGTGACATC TTAACTGAC CCGGGAGAAA TCACATTTAC CTGTGAATGT  
28001 TATTCTCAGT GCAACTTTCA TAAAAATATTG AAGGCAAAAT AAAGACATCT  
28051 CGGACATTAA AAAACAACCTG AGGCCAGCG TGGTGGCTCA CACCTGTAAT  
28101 ACCAGCACTT TGGGAAGCCA GACGAGGTCA AAAGTTCGAG ACCAGCCTGA  
28151 CCAACATGGT GAAACCCCAT CTTGACTAAA AATACAAAAA TTAGCCAGGC  
28201 ATGGTGGCGT GCACCTGTAA TCCCAGCTAC TCAGGAGGCA AAGGCAGGAG  
28251 AATTGCTTGA ACCTGGGAGG CAGAGGTTGC AATGAGCTGA GATCATGCCA  
28301 CTGCACTCCA GCCTGGGCAA CAGAGCAAGA CTCCATCTCG AAAAAATAAA  
28351 AATAAAAAAT GAACAACAAC AACAACTGAG AGTGTCTACC ACTCACTGGT  
28401 CTCCATCAAA GGGAGTTTGA AAGGGAGAAT TTTAGCAAGT CTATAGGAGA  
28451 CCGAGGTCGG GGGTATAAAT GATGTCAGCA AGGCAAACCTA CCTGCTGGCT  
28501 GTACAAACAT GAATGGTGCA GACTTTGATG GTGGGGGAAA GGTTCTGTTT  
28551 TAGTGACATAT TTAAAGAGTG CTAAATGCAC AAAAAACACA CTTGTGCCCCA  
28601 CACGCATGGG TTCATACGAA CAAGAATGCA CACAAAACAC ACTTGTGCCC  
28651 ACACGCATGG GTTCATACGA ACAAGAATGC ACACAAAACA CACTTGTGCC  
28701 CACACGCATG GGTTCATACG AACAGAATG CACACAAAAC ACACTTGTGC  
28751 CCACATGCAT GGGTTCATAT GAACAAAAGA TGCACGTTAA GCACGGGAGA  
28801 ATAGTGATCC ATGGAGAGGG AAGTGGAAG GGGTGGGCAG GGATAATAAA  
28851 TTTTCAAAAT ACATAAAGAA GCCTGTTAGA CAAAAGCATG ATTTACTTTT

28901 CATCAGTATA GATGAAAAGG CAGCCAACAT GAAAACAATT AAAAGCACAA  
 28951 TCACAAGAAG ACGAATGTTC ATGAATGTGA AGGCATACTG CATCGAAATA  
 29001 CAAGAAGCCA AAACAGACTG AAGAACAAAG GGGTCAAATA CACAGTTGGA  
 29051 AGGGGAGATG TCAGAACTTC TGCATTAGAA ACAGAGGGCG AGGGTGTTAA  
 29101 AAGTTCACAC CCATGTCGAA ATTTGGAACC AACATTGTAA AGTATACAAA  
 29151 GACACTGCCT CAAAAGGTAA ACAATTATAT TGTTATAATT GTTACAATGA  
 29201 ACAATTCTTA ATTGCTAATT CTTATTAAAT ACACATGGAA TAGTTTTAAA  
 29251 ATAGGCCACA TACCAGGCAA TATGGGGAAC TTCACAGAAC TAACGCAGCA  
 29301 GAGCCTGCCC TTTGTAGCTA CAATGCAATA AACTAGAATC AACAAATGAAC  
 29351 AGAGTGTTAC ATAAAGAGAA ATATAAACTT GAAGTCTGCA GCCTTAAACG  
 29401 CAACTATTAT ACAATAAAGA CGTCAGGACA GAAGTCCTAG CTGTGAAAAG  
 29451 TTCTCGGTCA TGAACACGCA CGAATGCCCG GCCAGCTCTG CGCTCCTTCG  
 29501 TCAGGACAGA AGTCCTCTCT GTGTGAAAAG TTCTCGGTCA TGAACACGCA  
 29551 CGAATGCCCG GCCAGCTCTG CAGTCCTTTG GTGGTCACCT CGCAAAGGAG  
 29601 AATTGATCGG AGCTGCTGCT TACGGGGTGA CACCCGTGGC AGCATCACTA  
 29651 CACGCACATC GTGACGGCAA CCGCGTGTGT CATGCAGCCC CCAACCGATG  
 29701 ATGAGAAACC AACCTGAACC TCACGTCAGA GGAAGGCTG TTATCTTTCC  
 29751 ATTCTCTTTA AGAAAAATTA TTTTAACACG AAGCAATAAT CAAGGGGTAG  
 29801 CAGCCAAGAG ATATAGAAAA AATAAGTTTG ACAGAGGTAT CCCAGGGAGT  
 29851 AACTAATCAA ATCTTGTTTT TTTCTTCCAT TTGATGATGT TTAGTGCACC  
 29901 TATCAGCTTT TTCATGTGCG AGCCTTGTTA TTGTTTTATC ACTTCAAATA  
 29951 AATATTTATT TTCATATCTA ATTTGGATTC ATATTGCTAT TATTATTTAA  
 30001 AGTAACCCCT AAAGTTTACA TATATTTTAA GCCCCACAAA ATCCATCTAC  
 30051 TCCTGCCTAT ATCTGTGAGT TTAATTTTCT TTTAAAATAA TCTGAATATA  
 30101 TACCAATGAG ATTTTCTTTT TTTTAATTTT TATATATATA TTTTTTTATT  
 30151 ATACTTTAAG TTCTAGGGTA CATGTGCACA ACGTGCAGGT TTGTTACATA  
 30201 TGTATACATG TGCCATGTTG GTGTCCTGCA CCAATTAAC~~CT~~ CGTCATTAC  
 30251 ATTAGGTATA TCTCCTAATG CTATCCCTCC CCCCPC~~CC~~ CACCCGATGA  
 30301 CAGGCCCCGG TGTGTGATGT TCCCCTTCCT GTGTCCAAGT GTTCTCCTTG  
 30351 TTCAGTTCCC ACCTACGAGT GAGAACATGC GGTGTTTGGT TTTTTGTGCC

30401 TTGGGATAGT TTGCTGAACT AGAAATACCA TTTGATCCAG CAATCCCATT  
30451 ACTGGGTATA TACCCAAAGG ATTATAAATC ATGCTTCCAT AAAGACACAT  
30501 GCTCACGTAT GTTTATTGCA GCACTATTCA CAATAGCAAA GACTTGAAAC  
30551 CAACCCAAAT GTCCATCAAT GATAGACTGG ATTAAGAAAA TGTGGCACAT  
30601 ATACACCATG GAATACTACG CAGCCATAGA AAACGATGAG TTCATGTCCT  
30651 TTGTAGGGAC ACGGATGAAG CTGGAAACCA TCATTCTCAC CAACAATGAG  
30701 ATTTTCTAAT GCCGGGCATG GATGATAGGT GTTTTGTTC TTGGCTTGCT  
30751 CCCTGGGTGC TGAAATAGTC AGGATTAGGA CAGCGGACCA TGGAGAAAGT  
30801 GCCCAGATGT GCTGGTACCT GCAGCTCAGG ACAAAGGGAG GCAGCGTGAA  
30851 TGACACGTTT TTTTATTTGG GTTCTTGCTG CTCCCACGTT ACTGTCACAA  
30901 TATTTTCTGT GGAGTTGAAA AGCAAATGCC CCAAGGCAAG CACAGAGGGT  
30951 GCAGGGGTTG GGGTGTCCCG AGGGCGGGG TGTGGGCGGG CCTGGACCCC  
31001 CTGCACTGCG GCTCCTCCTC TGTCCCTGGG GTTCCGTCCT CCTTGGTGGC  
31051 CTGCAGGGCC TCTGTCCACT GCCGGGCCAC CAACTAGGCT ACAGAGACAC  
31101 AGTGACGGTC AGGTTTCTCA GGTCCCCCG GCTAAGGAAC ACGTGAGCCT  
31151 TTCAGCCACC ACGTTGACCA CACATGCGAC AAAGGCAGAG GGGCCGCCTG  
31201 GTCACGGAGG ACGAGCAGCC CTTTGCCCCG TCTCCTGGAG GCTGGCGCTC  
31251 AAATGAGTAA ACCCAGGGAG TCCTGAGATT CTTCTGCTC AGACACAATG  
31301 CCCTAGTGAG GTCGCTTCCT GACACGAGTT TAAAGGATGC ACACGTGTTA  
31351 TAAATCTTAG AAAGCTGGAA GAGGGGAGAA AATCACTTCC AATCCCACCT  
31401 CCCCAGTGAC TCACTCGCGA GTTGGAACAG TTTCTTCCT GTCTTTTTTC  
31451 TAACAGAGCA TCTTACGTGG ATGCAGTGAC AAAGATGCGT CATCCGCCTG  
31501 CGACCGAAGG AAGCTGCGGG AAGAACACCC AAGGGCCTTG TGGCATCTCT  
31551 CCTGCGCCCT GCACTTCCCT GGGGCCAATG CCAAGGGCAA GGCCAGCCGT  
31601 GGCGGGCAGC AGCCCCCTCTG CTGGACACGG CGCCTGCCCC AAGGCCAAGG  
31651 CTCCAAACTC GGCCCCCTAC TGAGCATCTG ATTCCTGAG TCCCCCTCCG  
31701 AGGGCTCCAG GATTTGACCT CACATTTCCC GCCCGCTAA GTCTGTGTGC  
31751 TTTTAGGCCT GTGTGGCCTT CCCTGTCTGT GCCCTCAGTC TACCTGGACG  
31801 CTGCCATCCT TCAAAACCCA GCCCAGTCCC ACCTCCTGAA GGATGAAGGG  
31851 GCCCTGCCCC CCTGGCCTCC AGGGCGTCTT GGGCCAGGGC AGCCGCAGTG

31901 GTCCCTGCAT GTCCCCGTCA CAGAGGGTGC CTGTCCTGAA GGAGTGCAGT  
31951 CATCCCAACA GGCCTCGGCC CTCTCTGGCC ACCGAGGCTC CCTCGCCCTT  
32001 CCGGGTTGGG CTCAGCAGAG GCTCTTGGGG CTCAACTGGA GTGAACCGAG  
32051 TCGCGGATGG TGAAGTGAG GGGTACTTGG CCATGGAGGC CCAGGGCTTG  
32101 GCAGGGGGGC CCAGGGCGCA TTCCTGGAGG GGCCGCTTCC CGGGAAAGGT  
32151 CGCGCTGTGT TTGTGTTCCG GCTCTTTCCT CCGCCTGAGT TCCTCCGAGG  
32201 CCCTGGCTTA TCTGAACACA GGACCCTGTG TGCCACGGCA TGCACCATGT  
32251 CACGGTGAAT GTCACACAAA CAGGAACGTG CCTGGAGGCC GCCTGGACCT  
32301 GTCCTCCCGC CCATCAGCTG CCAAGGTGAT GCTGTGGACA TGGCCGTCCC  
32351 CACTTGCTTC CCCACTTATG GAGCTCCTCA TGGGCTCAAC ACCCCCTTCC  
32401 AGAAGGCTGG CCCCAGGAGG GGAGTCGCCG GCATCTGGAA GCCTGGCAGG  
32451 CCTCCGCAGC GTGGCCAGCA TGGGGTCCAT GAGTCAGGGG GTGGAAGAGG  
32501 CCCCCTGACA TGGGGAACAC ACTAAGCTCC TAGCCCAGCT CTTCTCCTGG  
32551 AGGACCGGAT GCCCTCGCTC TGCCGCTGGG GAAGGAGTGC GCAGGGGGCT  
32601 TGAGGCCAGC AAAGGAACTG CCCCCCAAAT CTTACTCCGG GATCGGCACT  
32651 TTCCTGGAAG ACCCCAGGAG TAAGGAGAAG GAAGAGCAGC GGGGAGGGAG  
32701 GGAGTCCCAA GCCAGGTGCC ACAGAGATGG GGCAGCGTGA CCCTGTCCAG  
32751 GTCACCTCCT GGGAAACTGA GGCCCAGCCA GGCAGAACTG GGTCTCCTGG  
32801 TTCCAGCCAC CACCCACCAC TCAGGCAGCT GCCACCTGGG AGGCAGGAGA  
32851 TGTCTGAGGT GGGGTCTTCA TACCACGAGC CACATGGAGG GAGGCCACGA  
32901 TCTCCGTCCT CCCACCTGCA CACACACCCT CCCCAGGTG AGCACTAGGA  
32951 AGCCCTCTCT GGGGTGGGGG TGGATGCTAA TGTTCTGTGC AGCTCTTTGG  
33001 TTCAGTGAGT CCTTTTTTGGT GAAAGGGCTC AGAGATGCGG CTAAAGCCTG  
33051 TGCCAGCCTT GCACTCCCAT GAGGTTGCTG TTACCCTCAT GCCACTTTCT  
33101 TGGGAGAGGA AGGTGAGGCT GGAAATGGTA AGGGGCTGAG CTCACTCAGA  
33151 GGCTAGCACG ACCCCAGGCT CACACCTGCC AGCAATGTTA AGAGGCACTG  
33201 GGAGCTGAGA AGGGCTGAGG GTGCCACGCT GGGGTTGGCC AGGTCCAACC  
33251 TGCCCCCTCC TTTGAGGCCC TCTCCTTCCT TCCTGTGCTG GCCTTTGGAG  
33301 CCCACGGTTT CAGCATCCAT CACGGACCCC ACCGCACACC CAAGCGCTGC  
33351 TGAGGACTTT AACTCCTGCT TTGGAATTTT CAGTGCTTCT GATTATTCGT

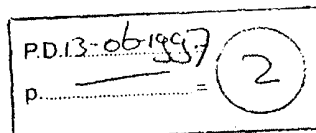
33401 AGACTCAAAA TCAGAAGAGG CTA CTTGCCC TAAACAGAAC ATGGTATATT  
33451 TGCATTGCAG ATCATTTTAC ATTTTAAAGA AAAAATGCAG TTTCTCTACA  
33501 AAGTCATCCC CCTCCAAAAA CAAAACAAAA CAAAAACCT CCCTAGGGAT  
33551 TTTGACAGGG CTGCCATCAG GCGTGTGACT GATTTGGGGA TAATCCCCC  
33601 ATCCCTTTGC TTTCCATTAC GCTGTGGCCC CCGCGTCCTT CAGGGCAGGT  
33651 GGCCTTGCTT TGTGAGCATC CTGCCCCTTT GGGGGCATT TCTGGGCACC  
33701 TGGTAGAGCC GTAGGTGAGA ATGGCCCCAG TGGTACTGGG CAGATGTGAG  
33751 CACAGGGACA ATGTTAGTGC AGGAAGTCAG GGGCCCGTCA GGGCTGTTGG  
33801 TGGCTGACGC GGCATCCATT CCCCTCCCCG CATCCCCCA ACACACCGGC  
33851 ACCCTCATGG ATGGTGACTG CACTCCCCAC TCCAGGGCTT GTCTGATTGA  
33901 TCTAATGGGG CGCCGTCCCT CTGGTCTGTG ATTGCTCTGG AATGAGCCAC  
33951 GGAGCCCAGA GAAGGAACAT GCAGAGGGCA TCTGAGAACT GCAGCAGGGG  
34001 ACCCAGTTGG GATGAGGCAG CCCTGGGGAC CTTGAGGAGA GAGGCAGGAA  
34051 GTGCCCAGTC CTTGACACCA TCCCCGAAGC CTGCCCCACG CCTGGACTCC  
34101 AGCCAAGGCA TGTAAGGCCA GGTGGACTCA GCAGTGTGAG GCAAAGCACG  
34151 GTGCCCCGGCT TGGATGCCAC ATGTGTTGAG CCAGCTGTGC TGTGGACAGG  
34201 GCGCTGTGGA CAGGGCACAG CAGACAGGCG GGGGATGTAG GCATCGGGAG  
34251 GGGGAGTCAT GGCTGGTGAG AGCACCCCTG GAGCGGGAAG TGCTGGGCAG  
34301 CGAGAGGTCA GGGAGGCATG GGGGCTTGGG ACGGCCACCA GCCACATCAG  
34351 AGCAGCACTG CTGCTGATTC CCAGCAACAT GGAATGAGGC CCCTTGGCAA  
34401 AGTAACAGAA CGTATCTTTG AAATGGCAGC AGCTGATCTT TAAAACAGAG  
34451 GCAAGCCGTT CCTGTCTCAA AACGTTTAAA TCCTCACCCC AAACCATTAT  
34501 GCGGCTCACT GCACCCGTGG AAAGCCAACA TCAATACCTG CGTGCATGCT  
34551 GGCCTCCCTG GCATGGGCAT TCTTTGAGGT TGCTGGCTCA AAGCCCCCAG  
34601 CACTGCCAGA GGGGGAAACC GGCAAACCA GGAATCCAA GGCTCAGCTC  
34651 AACAGCCACA CCCTTCCGTA GGCCAAAGGA AGCAAGAGCT TTTTCTTTCT  
34701 CACTGTTCTC AGGGATTCTT TGTGTTTTGTT TCTGTTTTTT GTTTTGTGTT  
34751 GTGTTTTTGGG AAGGGGACCA CGTCACAGGT TACTATAGAA ACAACGCCTG  
34801 AACTAGATGG ATGACGAAGC TGTTATTCCT CGGGGAGGCT CTTCAACTTG  
34851 AACCAATTGT GGGGCCATTC TTTAGAAAAC GGCTGTTTTT GCTCTGGGAA

34901 CATT TTTTAAA TTATTTCTTC TTCCTTTGCC TCAGATGGGA GGAGCATCGC  
34951 GCTACTGGAT CACCTCTAAT CATTTT TCC TTTTTTCTT TTTTTTATCC  
35001 AAAGCCCAGC TGGCCCATCA CAGGGTCATG TGTGAAAAGA GGGCTCCATG  
35051 CAGGTGTTCA GAAGAATGTT TTCAGTTTGT ACTCATCTCA GAAAATGCTT  
35101 CTGAAATGAT GTTAATAAAG AAAACAAGAT GCCAATGTGC ACATTCAGCC  
35151 TGGGCCTCAC GCAGCCTGGG GAAAATGTGC CAAAATGTCA GCATGGAAAAG  
35201 GATTCACAGT GATCTCTAAT TTCTTTCTCA GACCTTTATC AAGTTCTTAC  
35251 TATTTATGTG TTTGTTTATG GAGATGGGGT CTCACTTTGT TACCCAGGCT  
35301 GGTCTTGAAC TCCTGGGCTC AAGCGATCCT CCCGCCTCCA CCTCCTAAAG  
35351 CGTTGGGATT AAAGGTGTGA GCCTCCATGC CGGGCATTAC TTTTAAAGAG  
35401 CATATATAAC TTAGCCACGA GAAAAACAGC AGCATTTTTC AGCAACAACA  
35451 AAAGGGAGAG AGATGAGAGA AAGGCCCCCA GAGCTGGGCA GGGTTGGGGG  
35501 AAGAACCGTG GGCAGTGGGA GCAGCGACCC CAAATTCAGC CCCGGGCCCC  
35551 GCCTCCAGGC CAGCTCCTGG CAGGGTGGCC CCACCAGGGT GAAGGGCGGC  
35601 TGTCCGCGGC CTCCAGGTCA CCAGCAGGGC CTGGCTTATG ATCTATGAAC  
35651 CCCC GGCCGA TCTAGAGGTT AGGGCTCTAG ATGGGGCGGG GATT CGGTTA  
35701 GAGAGGGCAG GGTGGAGGGG AGGATAGACA GGAACCAGGG AGGTCCAGGC  
35751 CCAGTGGGGG CTGAGCTGAG CGGGGGAGGC GGGGCCGAGA AGGGCGGGGC  
35801 AGAGGAGTGG GATGAGCGTG GTTGGGGCAG GGCCAGTGGG GCAGGGTCAA  
35851 GGCTGTGTTG AGGGGTAGAG TGGGCGTGGT CGGAGCGGAG CCGGTGGGCG  
35901 GGACCCGGCG GTGTGGAGCT GCACCCGGAG TTTGTAAGGA TCAGGAAAGC  
35951 TGCCTGGAGC TGGGCAGACA GGCTGACCCA GGGACCCTGA CCACCTACCT  
36001 GAGCTTCCAC CCAGCTTCAA GGCTGTGGGG ATCGTTTCTT AGCCAAGGTG  
36051 AAAGGCCCTG AGCCCTCCAG TGGGCTCTGA GGGTTTGAAT GTTTATTCTA  
36101 AATGGGCTCC AACACGGAGT GCGCGTGCGC TCAAGGACCA GCTGGCTAGG  
36151 AGCGCCTTCT CTCCCCGAGG GGATCTGGAA TGTTCATCCG CGGTCTTCCC  
36201 AGTGGTCCTC TCCTGAGCAC CCTGGGCATG GGGTTGGGAA GTGGGCACAC  
36251 TTCTTCAGTG CGTGGTACTT TCTGTCCCTC CCAATGACAT TGGTCTCTTC  
36301 CAAATTCCCC AGGCATAGGA ACAGCCACAA GAGCCAACTC ACATTTCCCCA  
36351 AAGCCCTCTG GGGGCACGGC AGCGCCCTGG GCAGAGGAGA GAACTGCCCA

36401 GCTCAGCGGG GAGAAATCCC AGTGGTGCTT TCCGTTCCCC CCAGGCTCGA  
36451 TGA CTGCCAT ACTTACCAGC TGGGGGCGCT GGTGAGCACT GCCAGTCTCC  
36501 AGCGCCCGCC CTTACCCCGG GGCAGTGCTG CTTACAGCAC GGC GACTCCT  
36551 AAAGGCAAGG AGTTAAAGGC ACACGGATCC TAGCACTTTG GGGAGGCTGA  
36601 GGTGGGCGGA TC



XP-00207 80



ID HS1255458 standard; RNA; EST; 415 BP.  
XX  
AC AA460385;  
XX  
NI g2185598  
XX  
DT 13-JUN-1997 (Rel. 52, Created)  
DT 13-JUN-1997 (Rel. 52, Last updated, Version 1)  
XX  
DE zx51e06.r1 Soares testis NHT Homo sapiens cDNA clone 795778 5'  
DE similar to TR:G207551 G207551 ODORANT-BINDING PROTEIN. ;.  
XX  
KW EST.  
XX  
OS Homo sapiens (human)  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria; Primates;  
OC Catarrhini; Hominidae; Homo.  
XX  
RN [1]  
RP 1-415  
RA Hillier L., Allen M., Bowles L., Dubuque T., Geisel G., Jost S.,  
RA Kucaba T., Lacy M., Le N., Lennon G., Marra M., Martin J., Moore B.,  
RA Schellenberg K., Steptoe M., Tan F., Theising B., White Y., Wylie T.,  
RA Waterston R., Wilson R.;  
RT "WashU-Merck EST Project 1997";  
RL Unpublished.  
XX  
CC Contact: Wilson RK WashU-Merck EST Project Washington University  
CC School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis,  
CC MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email:  
CC est@watson.wustl.edu This clone is available royalty-free through  
CC LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for  
CC further information. Putative full length read The vector to vector  
CC length is Seq primer: -28m13 rev2 ET from Amersham.  
XX  
FH Key Location/Qualifiers  
FH  
FT source 1. .415  
FT /organism="Homo sapiens"  
FT /note="Vector: pT7T3D-Pac (Pharmacia) with a modified  
FT polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA  
FT was prepared from mRNA obtained from Clontech Laboratories,  
FT  
FT Inc., and primed with a Not I - oligo(dT) primer [5'  
FT TGTTACCAATCTGAAGTGGGAGCGGCCGCCCAATTTTTTTTTTTTTTTTTT 3'].  
FT Double-stranded cDNA was ligated to Eco RI adaptors  
FT (Pharmacia), digested with Not I and cloned into the Not I  
FT and Eco RI sites of the modified pT7T3 vector. Library went  
FT  
FT through one round of normalization to Cot5, and was  
FT constructed by Bento Soares and M. Fatima Bonaldo."  
FT /sex="male"  
FT /clone="795778"  
FT /clone\_lib="Soares testis NHT"  
FT /lab\_host="DH10B"  
FT mRNA <1. .>415  
XX  
SQ Sequence 415 BP; 106 A; 134 C; 108 G; 67 T; 0 other;

Hs1255458 Length: 415 September 30, 1998 10:31 Type: N Check: 5505 ..

1 AAGAAAACCT GATGCGGAAG ACGGAGGAGC CTGGCAAATA CAGCGCCTAT  
51 GGGGGCAGGA AGCTCATGTA CCTGCAGGAG CTGCCCAGGA GGGACCACTA  
101 CATCTTTTAC TGCAAAGACC AGCACCATGG GGCCTGCTC CACATGGGAA  
151 AGCTTGTGGG TAGGAATTCT GATACCAACC GGGAGGCCCT GGAAGAATT  
201 AAGAAATTGG TGCAGCGCAA GGGACTCTCG GAGGAGGACA TTTTCACGCC  
251 CCTGCAGACG GGAAGCTGCG TTCCCGAACA CTAGGCAGCC CCCGGGTCTG  
301 CACCTCCAGA GCCCACCCTA CCACCAGACA CAGAGCCCGG ACCACCTGGA  
351 CCTACCCTCC AGCCATGACC CTTCCCTGCT CCCACCCACC TGA CTCAAA  
401 TAAAGTCCTT CTCCC